

Giant QCD K -factors beyond NLO

Mathieu Rubin, Gavin P. Salam and Sebastian Sapeta

LPTHE
UPMC Univ. Paris 6
CNRS UMR 7589
Paris, France

Abstract

Hadronic observables in Z+jet events can be subject to large NLO corrections at TeV scales, with K -factors that even reach values of order 50 in some cases. We develop a method, LoopSim, by which approximate NNLO predictions can be obtained for such observables, supplementing NLO Z+jet and NLO Z+2-jet results with a unitarity-based approximation for missing higher loop terms. We first test the method against known NNLO results for Drell-Yan lepton p_t spectra. We then show our approximate NNLO results for the Z+jet observables. Finally we examine whether the LoopSim method can provide useful information even in cases without giant K -factors, with results for observables in dijet events that can be compared to early LHC data.

Contents

1	Introduction	2
2	The LoopSim method	5
2.1	The tree-level pure glue case	6
2.1.1	Attribution of structure to events	7
2.1.2	Constructing virtual (loop) events	8
2.1.3	Some examples	10
2.2	Treatment of flavour within LoopSim	11
2.3	Merging NLO calculations and beyond	14
2.4	Expected precision of the method	15
3	The reference-observable method	16
4	Validation: comparison to DY at NNLO	17
5	Results for the Z+jet process	19
5.1	Validation at \bar{n} LO	20
5.2	Results at \bar{n} NLO	22
6	QCD jet events as a testing ground	24
7	Conclusions	28
A	Recoil procedure	30
A.1	A particle recombines with the beam	30
A.2	A particle recombines with another particle	31
B	The LoopSim method and incoming partons	32
C	Secondary emitters in LoopSim	34

1 Introduction

At CERN’s Large Hadron Collider (LHC), it is widely anticipated that signals of new physics, for example supersymmetry, may manifest themselves as large excesses of data compared to expected QCD and electroweak backgrounds at high momentum scales [1, 2, 3, 4, 5, 6, 7]. The estimation of these backgrounds will be one of the elements in ascertaining the presence of any new physics from such signals. Consequently, considerable effort is being invested across the particle physics community in the development of methods to understand and predict backgrounds (some of the issues involved are described nicely in ref. [8]).

Given the QCD methods that are available today, some of the best prospects for obtaining systematic, accurate predictions of backgrounds involve next-to-leading order (NLO) QCD calculations. By carrying out a systematic expansion in the strong coupling

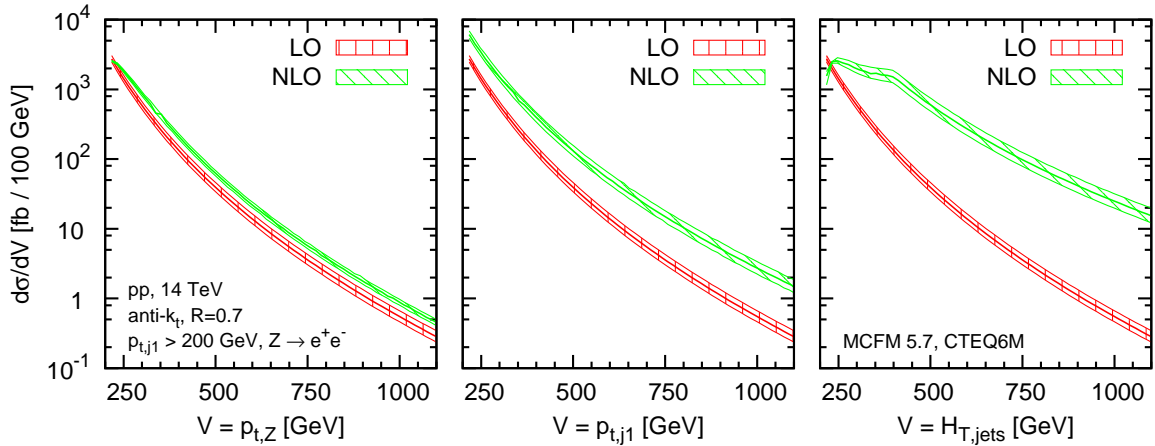


Figure 1: The LO and NLO distributions obtained with MCFM 5.7 [19] for three observables in Z +jet production: the Z transverse momentum (left), the p_t of the hardest jet (middle), and the scalar sum of the transverse momenta of all the jets, $H_{T,\text{jets}}$ (right). The bands correspond to the uncertainty from a simultaneous variation of $\mu_R = \mu_F$ by a factor of two either side of a default $\mu = \sqrt{p_{t,j1}^2 + m_Z^2}$. The jet algorithm is anti- k_t [20] with $R = 0.7$ and only events whose hardest jet passes a cut $p_t > 200$ GeV are accepted. The cross sections include the branching ratio $Z \rightarrow e^+e^-$.

and obtaining the first two terms (leading order (LO) and NLO) for a given process, one often obtains predictions that are accurate to 10 – 20%. The importance of NLO predictions in the LHC programme has motivated a large calculational effort destined to extend the range of processes known at NLO (for reviews, see refs. [9, 10]).

While the majority of NLO calculations show some degree of convergence relative to the LO results, several groups have commented in recent years on the appearance of K factors, ratios of NLO to LO results, that grow dramatically towards high transverse momenta [11, 12, 13, 14, 15, 16] (similar behaviour is visible also in [17, 18]). The problem generally occurs for hadronic observables (jet transverse momenta, etc.) in processes that involve heavy vector bosons or heavy quarks, at scales far above the boson or quark mass.

Fig. 1 illustrates this for the $pp \rightarrow Z$ +jet process at LHC (14 TeV) energies. It shows the distributions of three observables that are non-zero for configurations involving a Z -boson and one or more partons: the transverse-momentum of the Z -boson ($p_{t,Z}$), the transverse-momentum of the highest- p_t jet ($p_{t,j1}$) and the effective mass (scalar sum of the transverse momenta) of all jets ($H_{T,\text{jets}}$). At LO, all three distributions are identical. At NLO, the $p_{t,Z}$ observable is rather typical of a QCD observable: its distribution has a NLO K -factor of about 1.5, fairly independently of $p_{t,Z}$, and its scale dependence is reduced with respect to LO. The $p_{t,j1}$ distribution is more unusual: at high p_t it has a K -factor that grows noticeably with $p_{t,j1}$, reaching values of about 4 – 6, which is anomalously large for a QCD correction. The $H_{T,\text{jets}}$ observable is even more striking, with K -factors approaching 100.

Given that fig. 1 involves momentum scales where $\alpha_s \sim 0.1$, one is driven to ask how it is that such “giant” K -factors can arise. As touched on in [13], and discussed in more detail in [14, 15] for the $p_{t,j1}$ case, the answer lies in the appearance of diagrams with new

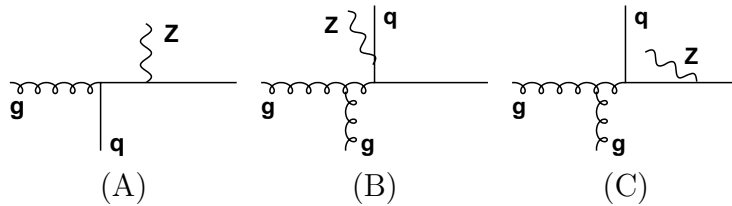


Figure 2: A) a LO contribution to Z+jet production; B) and C) two contributions that are NLO corrections to Z+jet observables but whose topology is that of a dijet event with additional radiation of a soft or collinear Z-boson either from a final-state quark (B) or an initial-state one (C).

kinematic topologies at NLO. This is illustrated in fig. 2: at LO the only event topology (A) is that of a Z-boson recoiling against a quark or gluon jet. One type of NLO diagram involves gluon radiation from this basic topology, giving modest corrections to all our observables. However, there are also NLO diagrams (B,C) whose topology is that of a dijet event, in which a soft or collinear Z-boson is radiated from outgoing or incoming legs. These diagrams do not contribute significantly to the $p_{t,Z}$ distribution, because the Z-boson carries only a moderate fraction of the total p_t . However when examining $p_{t,j1}$, it is irrelevant whether the Z boson is soft or not. Contributions B and C then lead to a result that is of order $\alpha_s^2 \alpha_{\text{EW}} \ln^2 p_{t,j1}/m_Z$, where the double logarithm comes from the integration over soft and collinear divergences for Z emission. The ratio of the NLO to LO results is therefore $\mathcal{O}(\alpha_s \ln^2 p_{t,j1}/m_Z)$,¹ rather than just $\mathcal{O}(\alpha_s)$, hence the K -factor that grows large with increasing p_t .² For the $H_{T,\text{jets}}$ observable the enhancement is even bigger because the dijet topology leads to $H_{T,\text{jets}} \sim 2p_{t,j1}$ instead of $H_{T,\text{jets}} = p_{t,j1}$ at LO.

While it is reassuring that we can understand the physical origins of the large K -factors in fig. 1, we are still left with doubts as to the accuracy of the NLO Z+jet predictions for $p_{t,j1}$ and $H_{T,\text{jets}}$, since they are dominated by the LO result for the Z+2-parton topologies. One way forward would be to calculate the full NNLO corrections for the Z+jet process. However, while work is progressing on NNLO calculations of $2 \rightarrow 2$ processes with QCD final states (see e.g. [22] and references therein), results are not yet available; nor are they likely to become available any time soon for some of the more complex processes where giant K -factors have been observed (e.g. some observables in $pp \rightarrow Wb\bar{b}$ [13, 17]). Alternatively one could simply try to avoid observables like $p_{t,j1}$ and $H_{T,\text{jets}}$ in inclusive event samples. For example, with additional cuts on the vector-boson momentum or a second jet, refs. [13, 15] showed that the K -factors are significantly reduced. However, given the many analyses that are foreseen at the LHC, it is likely that at least a few will end up probing regions where giant K -factors are present.

To understand how else one might address the problem of giant K -factors, one can observe that in our Z+jet example, the bottleneck in obtaining a NNLO prediction is the inclusion of the two-loop $2 \rightarrow Z + 1$ parton contributions and proper cancellation of all infrared and collinear divergences. Yet the two-loop (and squared one-loop) contribution

¹This differs from double electroweak (EW) logarithms, which involve terms like $\alpha_{\text{EW}} \ln^2 p_t/m_Z$, and are usually much smaller. Examples do exist of “giant” EW effects when tagging flavour [21].

²Part of the enhancement at high p_t also comes from the fact that one can have $qq \rightarrow qq$ scattering that emits a Z, whereas the qq partonic channel does not contribute at LO.

will have the topology of diagram A in fig. 2 and should not be responsible for the dominant part of the NNLO correction, which will instead come from diagrams with the topology of B and C, with either an extra QCD emission or a loop. So if one includes tree-level $2 \rightarrow Z + 3$ and 1-loop $2 \rightarrow Z + 2$ diagrams (i.e. $Z + 2$ jets at NLO) and supplements them with even a crude approximation to the two-loop $2 \rightarrow Z + 1$ result, one that suffices merely to cancel all divergences, then one should have a good approximation to the full NNLO result (a related observation has been exploited to obtain approximate NNLO results for high- p_t J/ψ production in [23]).

The purpose of this article is to develop a general method for obtaining such rough estimates of missing loop corrections. Our approach, called LoopSim, will be based on unitarity. After explaining how it works in section 2, and outlining a secondary “reference-observable” approach for control purposes, we will test the method by comparing its results to full NNLO predictions for lepton- p_t spectra in Drell-Yan production in section 4, apply it to our Z +jet observables in section 5 and finally, in section 6, examine whether it can be of use even in the absence of giant K -factors, specifically for a number of dijet observables.

2 The LoopSim method

The main ingredient of the LoopSim method is a procedure for taking a tree-level event with n final state particles and supplementing it with a series of events with $n - 1$ particles (approximate 1-loop events), $n - 2$ particles (approximate 2-loop events), etc., such that the sum of the weights of the full set of events is zero. This “unitarity” property will ensure that all the soft and collinear divergences of the tree-level matrix elements will cancel against identical divergences in the simulated loop contributions.

An outline of the procedure is given in fig. 3. Given a tree-level input event (a), the first step is to interpret it as a sequence of emissions (as if it had been produced by a parton shower), so that for example (diagram b) one can view particle 2 as having been emitted from particle 1, and particle 4 as emitted from the beam. The attribution of an emission sequence can be performed with the help of a suitable sequential-recombination jet algorithm and will be most meaningful in the limit that emissions are strongly ordered in angle and energy. The next stage is to decide which particles reflect the underlying hard structure of the event. If the event structure at the lowest possible order is that of a $2 \rightarrow 2$ scattering, then one should identify two outgoing “Born” particles. The Born particles will remain present in all the approximate “loop” events that are generated. They are represented as thick red lines in diagram (c). Again this step is most meaningful when all non-Born emissions are soft and collinear.

One then generates a set of simulated “1-loop” events by finding all ways of recombining one emitted particle with its emitter, diagrams (d,e). Each such “1-loop” event comes with a relative weight of -1 compared to the tree-level diagram. Similarly the set of simulated “2-loop” events is obtained by finding all ways of recombining two emitted particles with their emitter(s) (diagram f), each with relative weight $+1$; and so forth down to events where only Born particles remain (in fig. 3 this is already reached at the two-loop level). Note that the loop-diagrams drawn in fig. 3 are not intended to represent

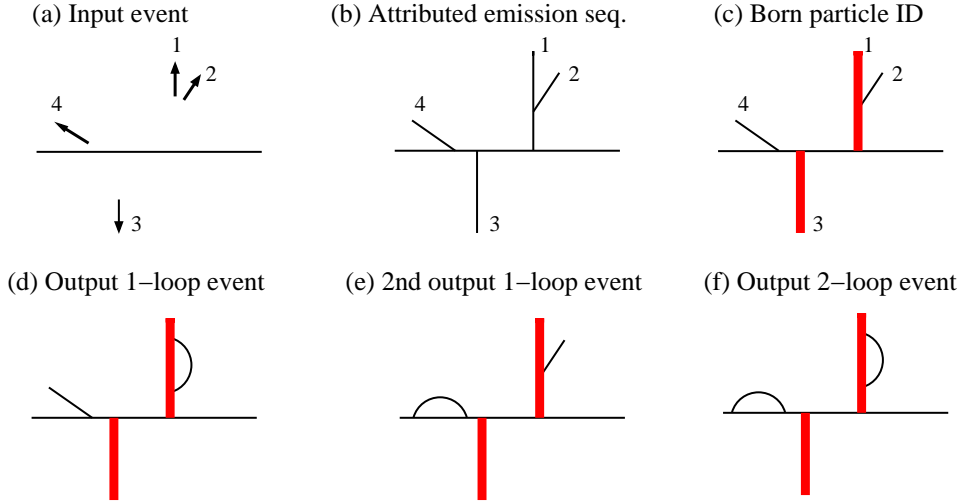


Figure 3: Sketch of the LoopSim procedure as applied to a tree-level event (a) with 4 outgoing particles (numbered) and the beam (horizontal line); diagram (b) shows the attribution of the emission sequence, (c) the identification of the Born particles (thick red lines), and (d)-(f) the resulting “looped” diagrams. These diagrams are relevant in approximating next-to-next-to-leading corrections to a process whose LO contribution has a $2 \rightarrow 2$ structure.

the actual Feynman diagrams that would be relevant at 1 and 2-loop level. Instead they indicate the way in which we have approximated the loop divergences, as the unitarising counterparts of the divergences that appear for each emission in the soft and collinear limits.

Given the above procedure for unitarising tree-level events, we shall see that it is then straightforward to extend it to event sets that also include exact loop diagrams.

2.1 The tree-level pure glue case

We start by examining the LoopSim procedure in the simple case of purely gluonic tree-level events. This will suffice to introduce most of the relevant concepts. Section 2.2 will then discuss some of the additional issues that arise for events with quarks and vector bosons, while the handling of events sets that include exact loop diagrams will be left to section 2.3.

It is helpful to introduce some notation: Firstly, b is the number of final-state particles present in the lowest relevant order (i.e. the number of final-state “Born” particles). For instance $b = 2$ if considering higher-order corrections to dijet events, as in fig. 3. E_n represents a generic event with n final state particles. So the starting event of fig. 3 would be labelled E_4 . Finally, U_l^b will be an operator that acts on an event E_n and returns all the events at l loops obtained from E_n using the LoopSim method. For instance, fig. 3d,e represents the action of $U_{l=1}^{b=2}$ on the input E_4 event (a).

The central part of the LoopSim method involves the construction of the operator U_l^b acting on E_n for all $l = 0 \dots n - b$ ($l \leq n - b$ because the number of real final state particles cannot be smaller than that of the lowest order event).

2.1.1 Attribution of structure to events

Recall that the primary function of the LoopSim method is to cancel the divergences that appear in the soft and collinear limits. In these limits, events can be interpreted as stemming from a sequence of probabilistic (parton-shower) type $1 \rightarrow 2$ splittings of some original hard Born particles. The knowledge of the splitting structure will help us generate loop events to cancel the divergences.

The attribution of a branching sequence is most easily performed using a sequential recombination jet algorithm (and is inspired by the CKKW matching procedure [24]). We will use the Cambridge/Aachen (C/A) algorithm [25, 26], which has important advantages over the k_t algorithm when dealing with nested collinear divergences (avoiding “junk” jets [25]).³

As a first step, to each of the $i = 1 \dots n$ particles in the event E_n , we assign a unique “identity” index $I_i \equiv i$.

We then run the Cambridge/Aachen (C/A) algorithm on the event. It repeatedly clusters the pair of particles that are closest in angle, i.e. with smallest $d_{ij} = \Delta R_{ij}^2 / R_{\text{LS}}^2$ where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ is the usual squared angular distance in the (y, ϕ) plane, and R_{LS} is a free parameter, the LoopSim radius. The C/A algorithm continues until all the $d_{ij} > 1$, at which point the remaining particles are deemed to cluster with the beam.

An $ij \rightarrow k$ clustering in the C/A algorithm can be reinterpreted as a $k \rightarrow ij$ splitting. The C/A algorithm does not distinguish in any way between i and j . However, in the soft limit, say $p_{tj} \ll p_{ti}$, rather than viewing k as splitting to i and j , it is a better reflection of the divergent structure of the amplitude to view k as having emitted a soft gluon j . Then i is nothing other than particle k with some small fraction of its energy removed. To account for this, in an $ij \rightarrow k$ clustering, if $p_{tj} < p_{ti}$, then we declare that the “identity” I_k of particle k should be the same as that of particle i , $I_k = I_i$. Also we record I_i as being a “secondary emitter” and remember that the object with identity I_j has been emitted from the object with identity I_i . (Exchange $i \leftrightarrow j$ if $p_{ti} < p_{tj}$). This is represented in fig. 3b by the fact that particle 1 is a straight line, off which particle 2 has been emitted; the identity of the $1 + 2$ combination is $I_{1+2} \equiv I_1 \equiv 1$. For an iB clustering, we record I_i as having been emitted from the beam.

The next step in attributing structure to the event is to decide which event particles should be viewed as the Born particles, i.e. which particles are responsible for the hard structure in the event. Inspired by the original formulation of the Cambridge algorithm [25], for every $ij \rightarrow k$ recombination we assign a k_t algorithm type hardness measure $h_{ij} = \min(p_{ti}^2, p_{tj}^2) \Delta R_{ij}^2 / R_{\text{LS}}^2$ [29, 30].⁴ For every beam recombination, we assign a hardness $h_{iB} = p_{ti}^2$.

We then work through the recombinations in order of decreasing hardness. For an $ij \rightarrow k$ recombination (or $k \rightarrow ij$ splitting), assuming i is harder than j , we mark $I_k \equiv I_i$ as a Born particle. If fewer than b particles have already been marked as Born particle, we also mark I_j as a Born particle. For an iB recombination, we mark I_i as a Born

³All jet clustering in this article is carried out using FastJet [27, 28].

⁴In the results shown later, we actually used $h_{ij} = \min(p_{ti}^2, p_{tj}^2) \Delta R_{ij}^2$, which however is identical for our default choice of $R_{\text{LS}} = 1$.

particle. This is repeated until b particles have been marked as Born (a particle may be marked more than once; in such a case its marking counts only once). As an example, in fig. 3, the hardest recombination will be between particle 3 and the beam, so particle 3 is marked as a Born particle. The next hardest recombination will that of $(1 + 2)$ with the beam. Therefore we mark $I_{(1+2)} = I_1 = 1$ as a Born particle. This exhausts the number ($b = 2$) of Born particles that need to be marked.

At the end of the above procedure, every particle will have been marked as emitted either from the beam or from another particle, and some particles will also have been marked as secondary emitters and/or Born particles. Thus in figure 3, particle 1 is labelled as having been emitted from the beam, it is a secondary emitter and a Born particle; particle 2 is labelled as having been emitted from particle 1; particle 3 is a Born particle, emitted from the beam; and particle 4 was emitted from the beam. The structure that we attribute is of course physically unambiguous only in the presence of strong ordering of emission angles and energies. However, as we shall argue in section 2.4, the mistakes that we make for non-ordered configurations should have a small impact for observables with giant K -factors.

2.1.2 Constructing virtual (loop) events

Once every particle is labelled in an event E_n , one can compute the result of $U_l^b(E_n)$, which is a set of events E_{n-l} . For an event E_n with respectively b Born particles and n_s non-Born secondary emitters, we define

$$v \equiv n - (b + n_s), \quad (1)$$

to be the maximum number of particles that will be allowed to become virtual in a given event. It is obvious that Born particles will not become virtual. Additionally, secondary emitters will also not become virtual. To understand why, consider the event



in which particle 3 is a secondary emitter, since it emitted 4. There is a divergence for 4 to be collinear to 3 only if 3 is a final-state particle. If instead 3 is made virtual, then the divergence for emitting 4 no longer exists (there is no divergence for emission from internal lines in a diagram), so that the weight for the diagram in which 3 is virtual would be the weight of the tree-level diagram times a small coefficient $\varepsilon \ll 1$. This simplest way of accounting for this is to approximate $\varepsilon = 0$ and thus not generate events in which secondary emitters are made virtual (a more detailed discussion is given in appendix C).

Having understood which particles can be made virtual, the operator U_l^b , when applied on an event E_n , generates all the $\binom{v}{l}$ diagrams in which l particles become virtual. For the virtual events to cancel the infrared and collinear divergences that appear in the tree-level diagram, we need an infrared and collinear (IRC) safe procedure to make particles virtual. For instance, the divergent weight of an event with two collinear partons i and j

has to be cancelled by that of corresponding virtual event (j makes a loop over i) when computing the distribution of any IRC safe observable; and two collinear partons, if not virtualised, have to remain collinear when another particle becomes virtual.

There are two ways for a particle j to make a loop:

- If it is labelled as clustering with particle i , then one has to spread the momentum of particle j over i and all the particles that are labelled as clustering with it but which were emitted after j according to the C/A clustering sequence (i.e. at smaller angle). The exact procedure is explained in detail in appendix A, and is designed to ensure that the recombination maintains any collinearity properties of non-looped particles and is invariant under longitudinal boosts. When j is the only particle that clusters with i , then the procedure becomes equivalent to adding the momenta of particles i and j , $p_k = p_i + p_j$, and then rescaling the momentum p_k such that its mass is set to 0, while leaving its transverse components p_x , p_y and its rapidity unchanged.
- If particle j is labelled as clustering with the beam, then when it is “looped” it is simply removed from the event. Note that looping particles with the beam is less trivial than it may seem at first sight, because of an interplay with factorisation and the PDFs. Nevertheless it can be shown, appendix B, that for particle types that are included in the PDFs it does make sense to loop them. A p_t imbalance will result from the looping of particles with the beam, and so after all loops have been made, we apply a transverse boosts to all remaining event particles, conserving their rapidities, so as to bring the total transverse momentum to zero (again, see appendix A).

There is some arbitrariness to our procedures for producing physical kinematics in the looped events. One avenue for future work would be to examine the impact of making different choices.

The operator U_l^b has the following properties

$$U_0^b = \mathbb{1}, \quad U_l^b(E_n) = 0 \quad \text{if } l > v. \quad (3)$$

If w_n is the weight of event E_n , then each of the events generated by the $U_l^b(E_n)$ operator has a weight

$$w_{n-l} = (-1)^l w_n. \quad (4)$$

Once all the $U_l^b(E_n)$ have been calculated for $l = 0 \dots n - b$, one has to combine them in order to subtract all the soft and collinear divergences that appear in the calculation of E_n and the virtual diagrams generated from it. This is done by the operator U_v^b , which is defined as

$$U_v^b \equiv \sum_{l=0}^v U_l^b. \quad (5)$$

It generates all the necessary looped configurations that have the same order in α_s as the original tree-level diagram. It is straightforward to see that the total weight of the

diagrams obtained from the U_{∇}^b operator is 0. Indeed, if we apply it to an event E_n whose maximum number of virtual particles is v , we get

$$w_n \sum_{l=0}^v (-1)^l \binom{v}{l} = 0, \quad (6)$$

for $v > 0$.

We note that the above procedure for approximating loop diagrams does not generate the finite terms needed to cancel the scale-dependence of lower-order diagrams. While it would be straightforward to include such terms, we believe that in the absence of full loop calculations, not including them helps ensure that the standard procedure of variation of renormalisation and factorisation scales is more likely to provide some form of reasonable estimate of the uncertainties on our results.

2.1.3 Some examples

In order to illustrate the action of the operator U_l^b , we give below some simple examples in the pure glue case. In each of these examples, only the Born particles are labelled with numbers

$$U_{l=1}^{b=2} \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = - \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array} - \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array}, \quad (7a)$$

$$U_{l=2}^{b=2} \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array}, \quad (7b)$$

$$U_{l=2}^{b=2} \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array} + \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array} + \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array}, \quad (7c)$$

$$U_{l=2}^{b=3} \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array}, \quad (7d)$$

$$U_{l=2}^{b=2} \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \begin{array}{c} \text{loop} \end{array}, \quad (7e)$$

$$U_{l=3}^{b=2} \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = 0. \quad (7f)$$

Eq. (7a) gives an example of singly-looped configurations (“1-loop diagrams”) generated by LoopSim when studying the $2 \rightarrow 4$ contributions to QCD dijet production. Eq. (7b) shows the “2-loop diagrams” generated from the same event. The next equation shows what happens if we add one more particle to the final state. If, eq. (7d), we now set the number of Born particles for the same event to be 3, we obtain only one 2-loop diagram instead of three, as represented in eq. (7d).⁵ Finally, the last two examples of eq. (7) give a case with a splitting: the emitter is not looped, even if it is not a Born particle.

We also give a few examples of the action of the U_V^b operator:

$$U_V^2 \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} - \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \text{ (with a loop on the right)} - \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \text{ (with a loop on the left)} + \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \text{ (with a loop on the top)} \quad (8a)$$

$$U_V^2 \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} - \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \text{ (with a loop on the right)} - \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \text{ (with a loop on the left)} + \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \text{ (with a loop on the top)} \quad (8b)$$

$$U_V^2 \left(\begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \right) = \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} - \begin{array}{c} 2 \\ | \\ \hline 1 \end{array} \text{ (with a loop on the right)}. \quad (8c)$$

In the last case, only one particle can become virtual because there are two secondary emitters which cannot be looped.

2.2 Treatment of flavour within LoopSim

Let us now examine some of the issues that arise if we are to extend the LoopSim method to processes with quarks and vector bosons.

We start with quarks and consider the situation depicted in fig. 4. In this case, applying the C/A algorithm as in the previous section will lead to the recombination of the two quarks q_1 and q_2 , which is clearly not physical. If flavour information is available

⁵One might reasonably be surprised by this: after all, the result for the exact two-loop diagrams is independent of any choice of number of Born particles. The point is that if one studies soft and collinear corrections to the 3-jet cross section, then for the events in eq. (7c) where the particle labelled 3 in eq. (7d) is virtual, the event will resemble a two-jet event and so not pass the 3-jet cuts. However if one studies the 3-jet cross section in a kinematic region where the cuts allow one of the jets to be much softer than the others, then to obtain sensible results it becomes necessary to use $b = 2$ and include all diagrams on the right-hand side of eq. (7c).

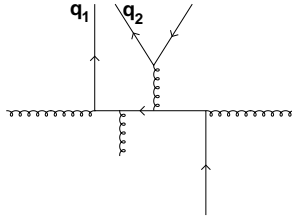


Figure 4: Example of an event where two quarks q_1 and q_2 may get recombined by the C/A algorithm.

for events, then one can veto on such a clustering, for instance by defining the clustering distance d_{qq} between two quarks to be infinite. As discussed in [31] such a modification alone is not sufficient to systematically guarantee sensible treatment of flavour in jet clustering. Refs. [31, 32] have both discussed the further modifications needed in the case of the k_t algorithm. A proper handling of flavour within LoopSim might seek to extend those modifications to the C/A algorithm. However, neither of the NLO programs that we use, MCFM and NLOJet++, provide information on particle flavours, so we defer such modifications to future work and just maintain the $d_{ij} = \Delta R_{ij}^2/R_{\text{LS}}^2$ distance for all partons. For observables that are not flavour-sensitive this should not be a major drawback, given the observation [31] that divergences associated with the mistreatment of flavour are strongly subleading. Were we to be interested in heavy tagged quarks, more careful treatment might well be needed. Note that there are also subtleties related to flavour and PDFs, discussed in appendix B.

What about non-QCD particles, specifically vector bosons? Let us examine the case of Z bosons. A Z can be emitted from quarks or antiquarks and we would like this to be reflected when establishing the approximate emission sequence, because if the Z has been emitted from a quark, then that quark is a secondary emitter and should not be looped. In other cases a Z boson may be the hardest isolated object in an event. Then it is to be considered a Born particle. On the other hand we won't necessarily wish to consider diagrams where a Z boson is looped, because they would represent electroweak corrections, not QCD corrections.

One issue in dealing with electroweak particles is that they are not emitted from gluons. If one could distinguish between quarks and gluons, then this could be accounted for during the C/A clustering, by defining the distance d_{gZ} between a Z and a gluon to be infinite. Since we will not know which partons are quarks or gluons, we adapt Frixione's isolation procedure [33] to decide if a Z boson relatively close in angle to a parton i is likely to have been emitted from i . More precisely, if

$$p_{ti} > \sqrt{p_{tZ}^2 + m_Z^2} \frac{\Delta R_{iZ}}{R_{\text{LS}}}, \quad (9)$$

then we define $d_{iZ} = \Delta R_{iZ}^2/R_{\text{LS}}^2$, otherwise $d_{iZ} = \infty$. When recombining i and Z into a particle k , then the identity index I_k is set equal to I_i (a quark and a Z give a quark). Our procedure means that a Z that is very collinear to a parton is always considered to be emitted from that parton — this makes sense because such configurations are much more likely to occur when the parton is a quark. In contrast a soft parton in the general

phenomenon of Bloch-Nordsieck violation [34] that is found when considering electroweak double logarithms. Since we in any case discard events in which electroweak bosons are looped, this should not be a problem for the practical use of LoopSim in events with W^\pm bosons.

2.3 Merging NLO calculations and beyond

Before explaining how we merge exact higher orders calculations, let us mention how we use the LoopSim method in practice on tree-level events at several different orders. We introduce the notation $X@{\bar{n}}^p\text{LO}$ to denote an approximation to the $N^p\text{LO}$ cross section for producing X , with all loop terms estimated through the LoopSim procedure. It is obtained by applying the U_{V}^b operator to all tree-level diagrams that can contribute up to $N^p\text{LO}$. For instance, one can write

$$Z@{\bar{n}}\text{LO} = U_{\text{V}}^1(Z@{\text{LO}}) + U_{\text{V}}^1(Z+j@{\text{LO}}) , \quad (13a)$$

$$Z+j@{\bar{n}}\text{LO} = U_{\text{V}}^2(Z+j@{\text{LO}}) + U_{\text{V}}^2(Z+2j@{\text{LO}}) , \quad (13b)$$

$$Z+j@{\bar{n}}\text{LO} = U_{\text{V}}^2(Z+j@{\text{LO}}) + U_{\text{V}}^2(Z+2j@{\text{LO}}) + U_{\text{V}}^2(Z+3j@{\text{LO}}) . \quad (13c)$$

Notice that $U_{\text{V}}^1(Z@{\text{LO}}) = Z@{\text{LO}}$ and $U_{\text{V}}^2(Z+j@{\text{LO}}) = Z+j@{\text{LO}}$. The terms $U_{\text{V}}^1(Z+j@{\text{LO}})$ and $U_{\text{V}}^2(Z+2j@{\text{LO}})$ simulate up to one-loop corrections, and $U_{\text{V}}^2(Z+3j@{\text{LO}})$ simulates up to two-loop corrections.

Now let us see how things work beyond tree-level accuracy. We define $E_{n,l}$ to be a generic event at l loops (exactly calculated) with n particles in the final state. We first consider the case where only one-loop corrections are computed exactly, so that we have tree-level events $E_{n,0}$ and exact one-loop events $E_{n-1,1}$. As before we can apply the unitarisation operator to the tree-level events, $U_{\text{V}}^b(E_{n,0})$. However, since we now include exact 1-loop contributions, $E_{n-1,1}$, we must remove the approximate 1-loop contributions $U_1^b(E_{n,0})$ that are contained in $U_{\text{V}}^b(E_{n,0})$. This alone is not sufficient, because among the extra contributions from the exact 1-loop terms, there will be pieces that are finite for a given $(n-1)$ -parton configuration, but that can lead to divergences when integrated over the $(n-1)$ -parton phase space. To cancel these extra divergences, we should introduce additional approximate higher-loop contributions, which can be obtained by applying the unitarisation operator U_{V}^b to the difference between the exact and approximate one-loop terms. So, rather than including just events $E_{n-1,1}$ and subtracting $U_1^b(E_{n,0})$, we include events $U_{\text{V}}^b(E_{n-1,1})$ and subtract $U_{\text{V}}^b(U_1^b(E_{n,0}))$. It is convenient to express this through a new operator $U_{\text{V},1}^b$ such that

$$U_{\text{V},1}^b(E_{n,0}) = U_{\text{V}}^b(E_{n,0}) - U_{\text{V}}^b(U_1^b(E_{n,0})) , \quad (14a)$$

$$U_{\text{V},1}^b(E_{n-1,1}) = U_{\text{V}}^b(E_{n-1,1}) , \quad (14b)$$

where the extra subscript 1 on the $U_{\text{V},1}^b$ indicates that it is the form to use when the exact 1-loop result is to be included. The action of $U_{\text{V},1}^b$ depends on the number of loops already included in the event on which it operates: we subtract the one-loop contribution returned by LoopSim only in tree-level events. With this notation, one can compute the

higher order corrections to eqs. (13) to one-loop accuracy,

$$Z@n\bar{n}\text{NLO} = Z@\text{NLO} + U_{\forall,1}^1(Z+j@\text{NLO}_{\text{only}}), \quad (15a)$$

$$Z+j@n\bar{n}\text{NLO} = Z+j@\text{NLO} + U_{\forall,1}^2(Z+2j@\text{NLO}_{\text{only}}), \quad (15b)$$

$$Z+j@n\bar{n}\text{NLO} = Z+j@\text{NLO} + U_{\forall,1}^2(Z+2j@\text{NLO}_{\text{only}}) + U_{\forall,1}^2(Z+3j@\text{NLO}_{\text{only}}), \quad (15c)$$

where the “only” subscript on $Z+nj@\text{NLO}_{\text{only}}$ means that we take the highest order that contributes, i.e. here $\alpha_s^{n+1}\alpha_{\text{EW}}$, since the LO, $\alpha_s^n\alpha_{\text{EW}}$, piece of $Z+nj@\text{NLO}$, is already taken into account in the $Z+(n-1)j@\text{NLO}$ contribution. This implies that one should use consistent renormalisation and factorisation scale choices across all different orders of the calculation. Note that in eq. (15) we have introduced the notation $\bar{n}^p\text{N}^q\text{LO}$ to denote an approximation to the N^{p+q}LO result in which the p highest loop contributions have been approximated with LoopSim.

The extension of the procedure beyond one-loop accuracy is simple. For instance, at two-loop accuracy, one has to subtract the approximated two-loop contribution $U_2^b(E_{n,0}) - U_1^b(U_1^b(E_{n,0}))$ in eq. (14a), and the other approximated two-loop contribution $U_1^b(E_{n,1})$ in eq. (14b), giving

$$U_{\forall,2}^b(E_{n,0}) = U_{\forall}^b(E_{n,0}) - U_{\forall}^b(U_1^b(E_{n,0})) - U_{\forall}^b[U_2^b(E_{n,0}) - U_1^b(U_1^b(E_{n,0}))], \quad (16a)$$

$$U_{\forall,2}^b(E_{n-1,1}) = U_{\forall}^b(E_{n-1,1}) - U_{\forall}^b(U_1^b(E_{n-1,1})), \quad (16b)$$

$$U_{\forall,2}^b(E_{n-2,2}) = U_{\forall}^b(E_{n-2,2}). \quad (16c)$$

Therefore, once $Z+j@n\bar{n}\text{NLO}$ is calculated, one may compute for instance

$$Z@n\bar{n}\text{NNLO} = Z@n\bar{n}\text{NLO} + U_{\forall,2}^1(Z+j@n\bar{n}\text{NLO}_{\text{only}}). \quad (17)$$

To be complete, let us mention the generalisation of our procedure to m -loop accuracy

$$U_{\forall,m}^b(E_{n-l,l}) = U_{\forall}^b(E_{n-l,l}) + \sum_{j=1}^{m-l} (-1)^j \sum_{\substack{l_1, \dots, l_j \geq 1 \\ l_1 + \dots + l_j \leq m-l}} U_{\forall}^b \circ U_{l_1}^b \circ \dots \circ U_{l_j}^b(E_{n-l,l}). \quad (18)$$

We noted at the end of section 2.1.2 that the plain LoopSim procedure does not generate the finite terms needed to cancel residual scale dependence from lower orders. With the introduction of the exact loop contributions, those finite terms do now get included. Thus for a given number of exact plus simulated loops, as we increase the number of exact loops, we should expect to see reductions in scale uncertainties.

2.4 Expected precision of the method

Let us briefly explain why the LoopSim method is expected to work in the presence of giant K -factors. We consider an observable A computed respectively at NLO and $\bar{n}\text{LO}$. We define $K_{\text{NLO}}^{(A)}$ such that

$$\sigma_{\text{NLO}}^{(A)} = K_{\text{NLO}}^{(A)} \sigma_{\text{LO}}^{(A)}, \quad (19)$$

and we assume that $K_{\text{NLO}}^{(A)} \gg 1$. This huge K -factor may come from logarithmic enhancements in the real NLO diagram or the appearance of new scattering channels in the

perturbative expansion. The computation of $\sigma_{\bar{n}\text{LO}}^{(A)}$ gives the exact real part of the NLO calculation as well as the divergent terms of the virtual correction. Therefore

$$\sigma_{\bar{n}\text{LO}}^{(A)} - \sigma_{\text{NLO}}^{(A)} = \mathcal{O}\left(\alpha_s \sigma_{\text{LO}}^{(A)}\right), \quad (20)$$

where, in writing $\mathcal{O}(\alpha_s \sigma_{\text{LO}}^{(A)})$, we mean that the term missing in the $\bar{n}\text{LO}$ calculation, the finite part of the 1-loop correction, is not especially enhanced. This leads to

$$\sigma_{\bar{n}\text{LO}}^{(A)} = \sigma_{\text{NLO}}^{(A)} \left(1 + \mathcal{O}\left(\frac{\alpha_s}{K_{\text{NLO}}^{(A)}}\right) \right). \quad (21)$$

The relative difference between the approximate and exact NLO calculations is thus suppressed by the inverse K -factor.

Next, consider $\bar{n}\text{NLO}$ accuracy. The difference between $\sigma_{\bar{n}\text{NLO}}^{(A)}$ and $\sigma_{\text{NNLO}}^{(A)}$ comes from the parts of the two-loop corrections that are finite and associated with the LO topology, so that they should be free of the enhancements that led to the large NLO K -factor. This implies

$$\sigma_{\bar{n}\text{NLO}}^{(A)} - \sigma_{\text{NNLO}}^{(A)} = \mathcal{O}\left(\alpha_s^2 \sigma_{\text{LO}}^{(A)}\right). \quad (22)$$

If we define $K_{\text{NNLO}}^{(A)}$ such that $\sigma_{\text{NNLO}}^{(A)} = K_{\text{NNLO}}^{(A)} \sigma_{\text{LO}}^{(A)}$, then we can write

$$\sigma_{\bar{n}\text{NLO}}^{(A)} = \sigma_{\text{NNLO}}^{(A)} \left(1 + \mathcal{O}\left(\frac{\alpha_s^2}{K_{\text{NNLO}}^{(A)}}\right) \right). \quad (23)$$

If $K_{\text{NLO}}^{(A)} \gg 1$, one can expect $K_{\text{NNLO}}^{(A)} \gg 1$ too.

3 The reference-observable method

Given the novelty of the LoopSim method, it is useful to have an alternative way of estimating the size of the NNLO contributions that we will approximate with LoopSim. Here we outline such an alternative method, which, though less flexible than the LoopSim approach, will provide a valuable cross-check and help us build our confidence in results of the LoopSim method.

Let us explain it for observables in the $Z+j$ process. Our aim is to estimate $\sigma_{\text{NNLO}}^{(A)}$ for some observable A .⁸ We assume that we have a reference observable which is identical to the observable A at LO. For instance, one might consider $\text{ref} = p_{t,Z}$ and $A = p_{t,j}$. We can write the NNLO $Z+j$ prediction for A in terms of the NNLO prediction for the reference observable plus the NLO $Z+2j$ difference between A and the reference cross section

$$\sigma_{Z+j@\text{NNLO}}^{(A)} = \sigma_{Z+j@\text{NNLO}}^{(\text{ref})} + (\sigma^{(A)} - \sigma^{(\text{ref})})_{Z+j@\text{NNLO}}, \quad (24a)$$

$$= \sigma_{Z+j@\text{NNLO}}^{(\text{ref})} + (\sigma^{(A)} - \sigma^{(\text{ref})})_{Z+2j@\text{NLO}}. \quad (24b)$$

⁸More precisely, $\sigma^{(A)}$ is the cross section for the observable A to pass some given cuts; it is only for brevity that we use here the somewhat inaccurate shorthand “cross section for observable A ”.

The second equality is possible because 2-loop NNLO corrections to $Z+j$ have the topology of $Z+j$ at LO. Therefore, their contributions to the observables A and ref are identical and cancel in the difference in eq. (24a).

If we have reason to believe that the perturbative expansion for the reference observable converges well, we can conclude that $\sigma_{Z+j@NNLO}^{(\text{ref})} - \sigma_{Z+j@NLO}^{(\text{ref})}$ is genuinely a small correction. Then

$$\sigma_{Z+j@NNLO}^{(A)} \simeq \sigma_{Z+j@NLO}^{(\text{ref})} + (\sigma^{(A)} - \sigma^{(\text{ref})})_{Z+2j@NLO}, \quad (25)$$

i.e. we approximate the NNLO distribution for A in terms of the NLO distribution for the ref observable and a NLO calculation for difference between the A and ref distributions, both of which are exactly calculable. The missing part is suppressed by a relative factor $1/K^{(A)}$, as for the LoopSim method. For $Z+j$, one can see from fig. 1 that $p_{t,Z}$ seems to be an acceptable reference observable for $p_{t,j}$ and $H_{T,\text{jets}}$.

In the sections that follow we shall, for brevity, refer to the RHS of eq. (25) as “ \bar{n} NLO” even though it does not quite adhere to our the meaning of \bar{n} NLO as set out in section 2, i.e. in terms of the specific sets of tree-level and loop diagrams that are included exactly.

4 Validation: comparison to DY at NNLO

The cross section for the Drell-Yan process is known with exclusive final states up to NNLO accuracy [35, 36]. Above a certain value of lepton transverse momentum, one finds giant corrections to the lepton p_t spectra when going from LO to NLO and large ones from NLO to NNLO. This gives us an opportunity to directly test the performance of the LoopSim method by comparing its \bar{n} NLO results to exact NNLO spectra for lepton pair production.

Before examining \bar{n} NLO results, it is useful to compare \bar{n} LO with NLO. If they are in reasonable agreement for some observable, then that serves as a first indication that the LoopSim estimate of missing loop corrections is sensible for that observable.

Fig. 5 gives the comparison of the \bar{n} LO, NLO and LO results for the production of an e^+e^- pair within the mass window of $66 < m_{e^+e^-} < 116$ GeV at a proton-proton centre of mass energy of 14 TeV. The left-hand plot shows the cross section differential in the transverse momentum of the harder of the two leptons. The right-hand plot gives the corresponding K factor with respect to LO. The results were obtained with MCFM 5.3 [37, 19], with its default set of electroweak parameters and NNLO MSTW2008 parton distribution functions. The uncertainty bands in Fig. 5 correspond to varying the renormalisation and factorisation scales $\mu_R = \mu_F$ by a factor of $\frac{1}{2}$ and 2 around a default choice of m_Z . In the \bar{n} LO result we fixed the value of the LoopSim radius parameter to be $R_{LS} = 1$, which naturally places interparticle and particle-beam clustering on the same footing (though the \bar{n} LO result here is actually independent of R_{LS} , because there is at most one isolated QCD parton in the final state).

There are three relevant regions of transverse momentum in fig. 5. For $p_{t,\text{max}} \lesssim \frac{1}{2}m_Z$ (low p_t) the distribution is dominated by on-shell Z-bosons and its shape is governed by the angular distribution of the Z decays in their centre-of-mass frame. The peak close

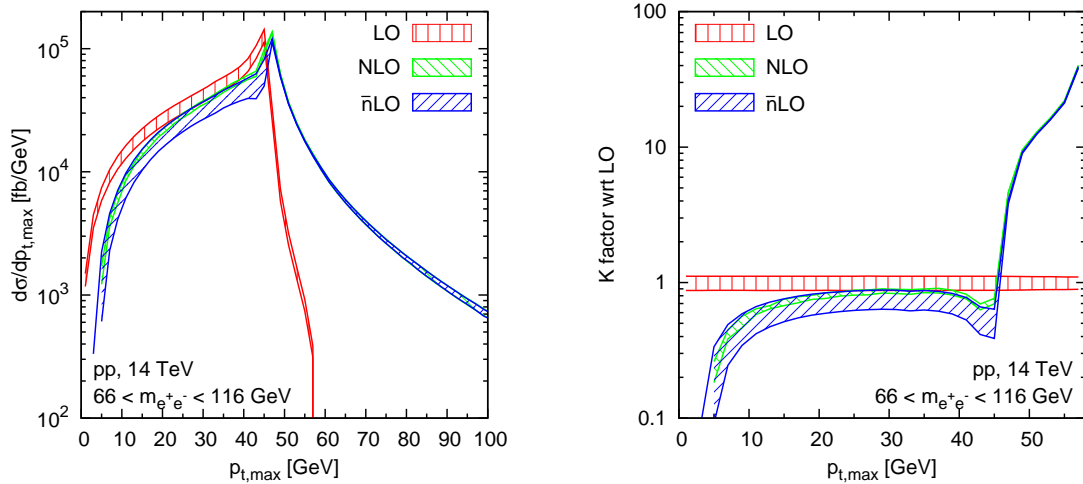


Figure 5: Comparison between \bar{n} LO results from LoopSim (with $R_{LS} = 1$) and exact NLO results for the Drell-Yan process. The left-hand plot shows the transverse momentum spectrum of the harder lepton, while the right-hand plot gives the corresponding K factors w.r.t. LO. The uncertainty bands were obtained by varying $\mu_R = \mu_F$ by a factor of $\frac{1}{2}$ and 2 around a default choice of m_Z .

to $\frac{1}{2}m_Z$ corresponds to Z-bosons that decay in a plane at right-angles to the beam. For $\frac{1}{2}m_Z \lesssim p_{t,max} < 58 \text{ GeV}$ (intermediate p_t), the LO distribution comes from Z-bosons that are off shell, which allows the p_t of the lepton to be larger than $\frac{1}{2}m_Z$. The narrow width of the Z causes the distribution to fall very steeply. The 58 GeV upper edge of this region is a consequence of our cut on $m_{e^+e^-} < 116 \text{ GeV}$. Above 58 GeV (high p_t) the LO distribution is zero.

In the low p_t region, the NLO correction is moderate and negative. There is no strong reason to believe that the LoopSim method should work here, but it turns out that the \bar{n} LO result reproduces the structure of the correction, even if its scale dependence remains much larger than that of the NLO result (this is because the LoopSim procedure does not include the finite terms that would partially cancel the LO scale dependence). In the intermediate p_t region, we see a “giant” NLO K -factor. It comes about because initial-state radiation can give a boost to the Z-boson, causing one of the leptons to shift to higher p_t (it becomes the “max” lepton). The spectrum of QCD radiation falls much less steeply than the Z-boson lineshape, so this NLO correction dominates over the LO result. In this region the exact loop correction, proportional to the LO result, becomes almost irrelevant and we see near perfect agreement between \bar{n} LO and NLO. In the high- p_t region only the real emission diagrams of Z@NLO contribute and \bar{n} LO becomes identical to NLO (both correspond to the Z+j@LO result). Similar results hold for the p_{t,e^\pm} distribution, while the $p_{t,min}$ lacks the giant K -factor in the intermediate region.

A similar comparison between \bar{n} NLO and NNLO spectra is shown in fig. 6. The NNLO results were obtained with DYNNLO 1.0 [36, 38, 39], used with a set of electroweak parameters compatible with that of MCFM.⁹

⁹ In its $\mathcal{O}(\alpha_s)$ and $\mathcal{O}(\alpha_s^2)$ contributions, DYNNLO includes among its parameters a cut on the p_t of

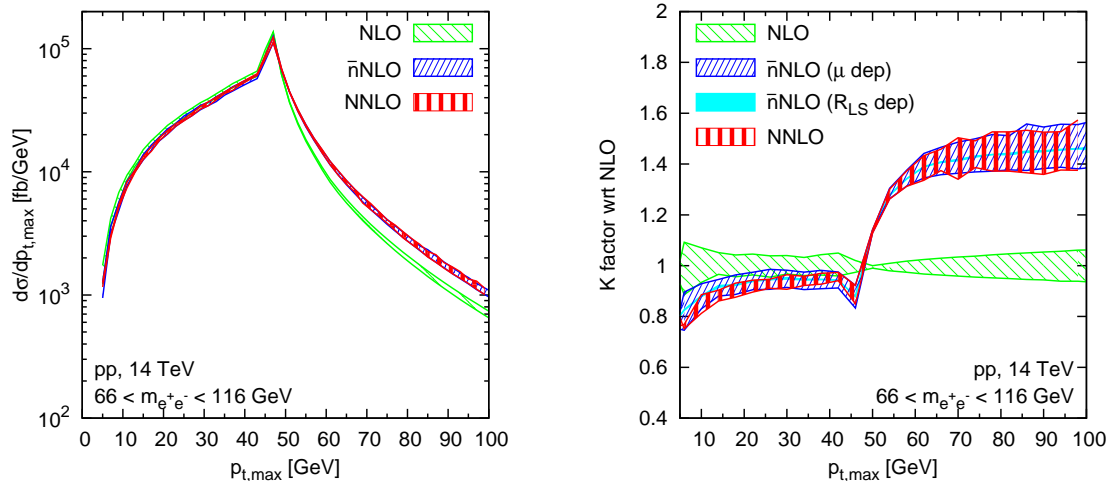


Figure 6: Comparison between \bar{n} NLO results from LoopSim+MCFM (with $R_{LS} = 1$) and full NNLO results for the Drell-Yan process from DYNNO. The left-hand plot shows the transverse momentum spectrum of the harder lepton, while the right-hand plot gives the corresponding K factors w.r.t. NLO. The uncertainty bands come from varying the factorisation and renormalisation scales by factors 1/2 and 2. In the right-hand plot we also show the (thin) band related to changing the \bar{n} NLO R_{LS} parameter from 0.5 to 1.5, at fixed $\mu_R = \mu_F = m_Z$.

In the low- p_t region we find quite good agreement between the \bar{n} NLO and NNLO results (with somewhat larger uncertainty bands for \bar{n} NLO). Such a result was not guaranteed a priori, even if it is not entirely surprising given the reasonable agreement that we saw between \bar{n} LO and NLO. In the intermediate p_t region, where the NNLO/NLO corrections are substantial, the agreement is excellent. This was expected. At high p_t the agreement should be exact, and does seem to be, within statistical fluctuations. The dependence on R_{LS} (shown in the right-hand plot) has been estimated by varying its value from 0.5 to 1.5. The effects are small.

Finally, we note that similar features and a similar level of agreement between \bar{n} NLO and NNLO are to be found in the $p_{t,\min}$ and p_{t,e^\pm} distributions.

5 Results for the Z+jet process

In the previous section, we studied the Z production process and showed that our procedure correctly reproduces the p_t distribution of the hardest lepton at NNLO, even, unexpectedly, in regions where the K -factor is not large. In this section we study the Z+j process, whose NNLO cross-section is not known yet, but which leads to giant K -

the Z boson. The cut is applied to both real and virtual terms and its impact should vanish as it is taken towards zero. It is, however, required to be non-zero for the numerical stability of the MCFM Z+j NLO calculation that is among the components of DYNNO. We set the cut equal to 0.1 GeV in the $\mathcal{O}(\alpha_s)$ term and to 1 GeV in the $\mathcal{O}(\alpha_s^2)$ term. A related 1 GeV cut was placed on the $\mathcal{O}(\alpha_s^2)$ piece of the \bar{n} NLO result (while none was used at $\mathcal{O}(\alpha_s)$). The impact of the 1 GeV cut is small but not entirely negligible close to the peak (where, physically, NNLO should in any case be supplemented with a resummation).

factors at NLO for some observables as explained in the introduction. Therefore, their NNLO contributions are expected to be accurately described by the LoopSim method. Throughout this section we use MCFM 5.7, including the Z+2j process at NLO [40], with the NLO CTEQ6M PDFs. We will take three different values for the renormalisation and factorisation scales: $\mu_R = \mu_F = \frac{1}{2}\mu_0$, μ_0 and $2\mu_0$, with

$$\mu_0 = \sqrt{m_Z^2 + p_{t,j1}^2}, \quad (26)$$

where $p_{t,j1}$ is the transverse momentum of the hardest jet. At high p_t , this scale choice should be quite similar to that used in [41] and has the same p_t scaling as those in [14, 15]. The R_{LS} uncertainty is measured at $\mu_R = \mu_F = \mu_0$ using three different values for it: $R_{LS} = 0.5, 1, 1.5$.

In addition to the 3 observables shown in the introduction, $p_{t,Z}$, $p_{t,j1}$ and $H_{T,jets} = \sum_{i=1}^{\infty} p_{t,ji}$, we will also consider

$$H_{T,tot} \equiv H_{T,jets} + p_{t,Z}. \quad (27)$$

We only include events for which $p_{t,j1} > 200$ GeV.

5.1 Validation at $\bar{n}LO$

As a first investigation of the performance of the LoopSim method, let us examine how the $\bar{n}LO$ approximation compares to the full NLO result. Fig. 7 shows the K -factors for the $\bar{n}LO$ and NLO predictions, with uncertainty bands from scale and R_{LS} variations.

In the upper-left plot, one sees that the $\bar{n}LO$ prediction for the $p_{t,Z}$ distribution gives a somewhat smaller K -factor than the NLO result. We interpret this as being because certain genuine loop effects are not taken into account by the LoopSim method, for example those related to threshold logarithms, which depend crucially on the factorisation scheme of the parton distribution functions. The $\bar{n}LO$ result does, however, reproduce the p_t dependence of the K -factor, i.e. the dip towards $p_t = 200$ GeV. This dip arises because of the requirement in our event selection that there should be at least one jet with $p_t > 200$ GeV. At LO this induces a step-function in the $p_{t,Z}$ distribution at 200 GeV. At NLO, soft and collinear emissions smoothen out that threshold and the $\bar{n}LO$ calculation correctly reproduces the resulting interplay between real and virtual terms.

In the three remaining plots of fig. 7, for $p_{t,j1}$, $H_{T,jets}$ and $H_{T,tot}$, all of which have giant K -factors, one sees good agreement between the $\bar{n}LO$ and NLO results. This is because the dominant NLO contribution comes from events in the B and C-type configurations of fig. 2, for which there is no corresponding QCD loop correction. The LoopSim method merely serves to cancel the divergences that arise from soft and collinear emissions off A-type configurations and these are not dominant overall.

The R_{LS} dependence, also shown on these four plots, only comes from 1-loop events generated by LoopSim. Therefore, for an observable A studied in Z+j@ $\bar{n}LO$ with two different values R_0 and R_1 for R_{LS} , one can write:

$$\sigma_{Z+j@nLO,R_1}^{(A)} - \sigma_{Z+j@nLO,R_0}^{(A)} = \sigma_{Z+j@nLO,R_1}^{(p_{t,Z})} - \sigma_{Z+j@nLO,R_0}^{(p_{t,Z})} \quad (28)$$

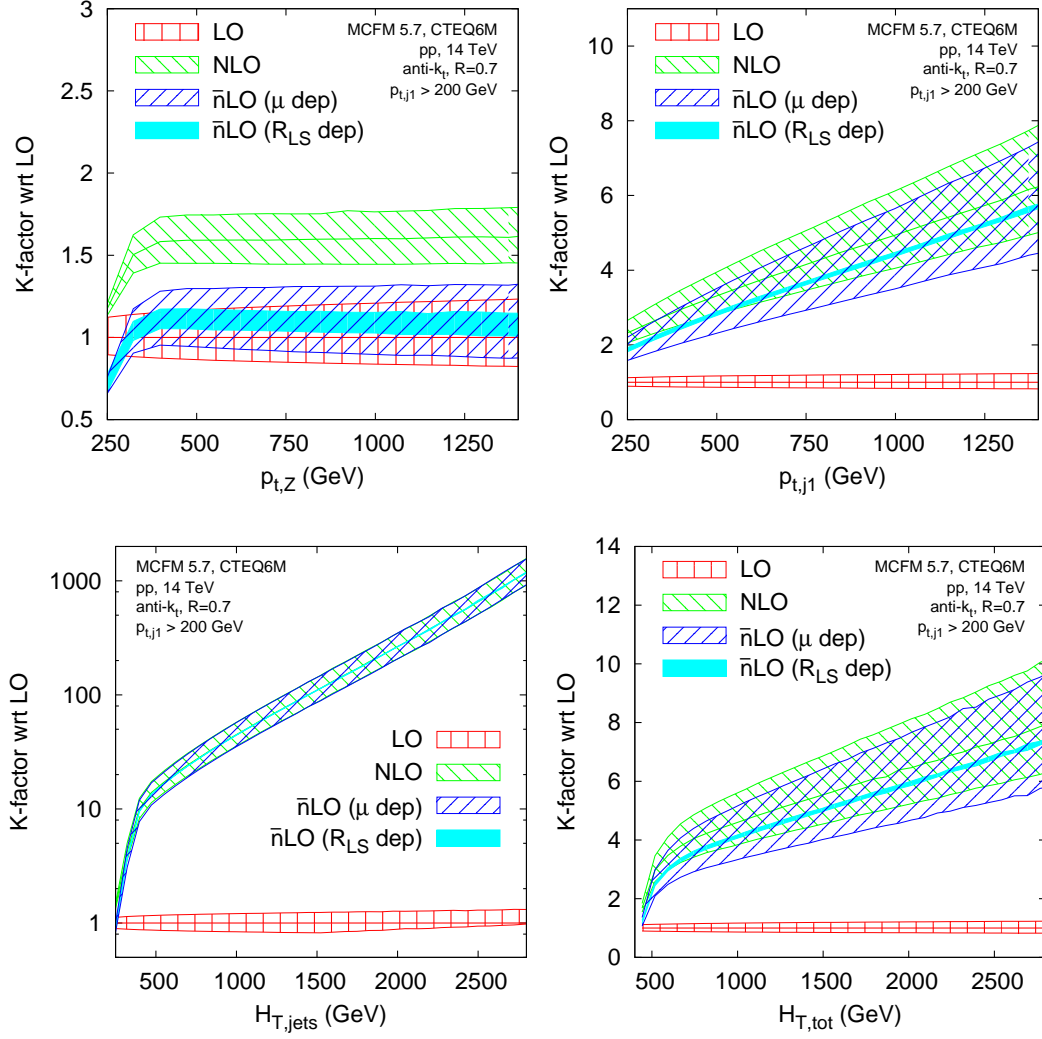


Figure 7: Comparison of the \bar{n} LO/LO K -factor with the NLO/LO K -factor, together with their scale and R_{LS} uncertainties for four observables in the Z+jet process.

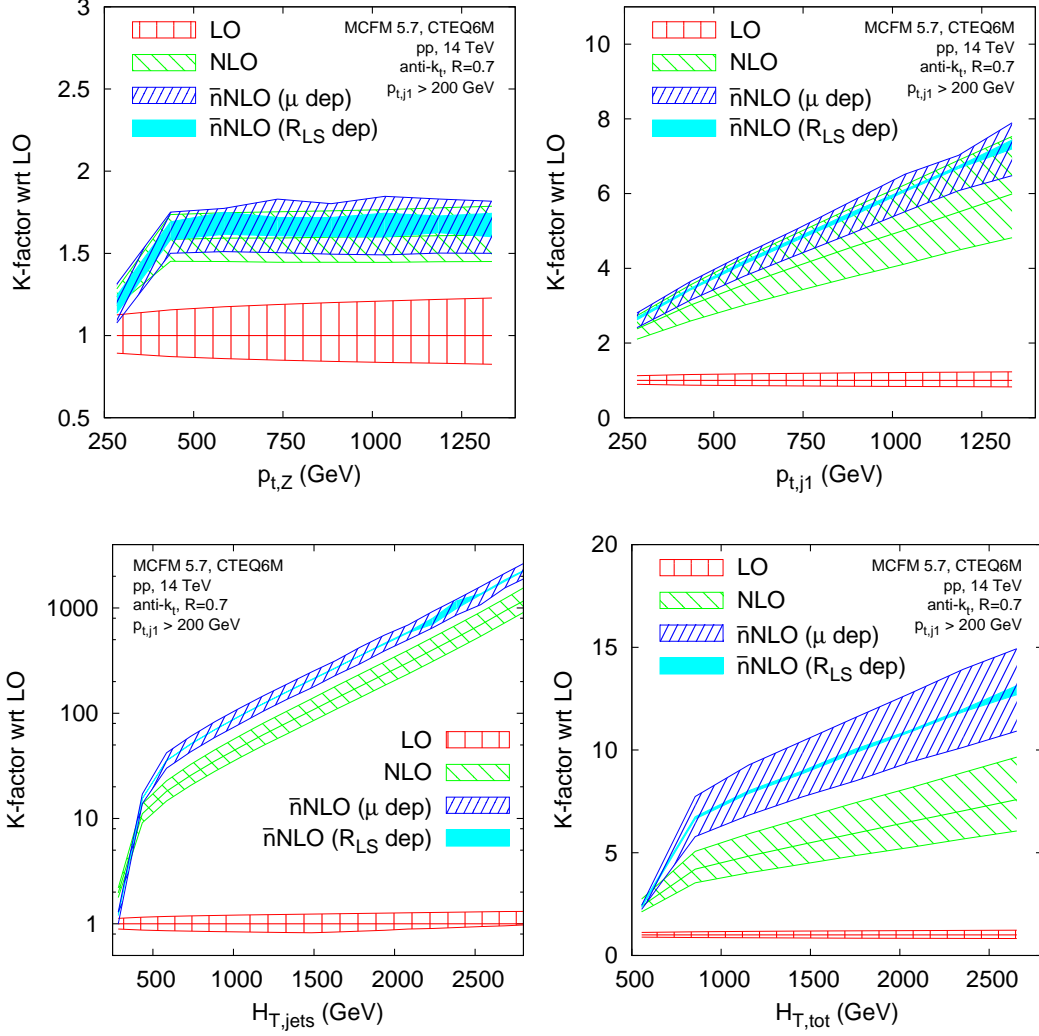


Figure 8: Comparison of the \bar{n} NLO/LO K -factor with the NLO/LO K -factor, together with their scale and R_{LS} uncertainties for four observables in the Z +jet process

as long as A coincides with $p_{t,Z}$ at LO (it does for each of $p_{t,j1}$, $H_{T,jets}$ and $\frac{1}{2}H_{T,tot}$). This means that the *absolute* uncertainty due to R_{LS} is the same for A and $p_{t,Z}$. Therefore, the *relative* uncertainty due to R_{LS} is expected to be roughly inversely proportional to the K -factor for A , in analogy with the discussion of sec. 2.4. This explains why the R_{LS} dependence (solid cyan band) looks significantly smaller for $p_{t,j1}$, $H_{T,jets}$ and $H_{T,tot}$ than it does for $p_{t,Z}$ plot.

5.2 Results at \bar{n} NLO

Results at \bar{n} NLO are given in fig. 8. In the case of $p_{t,Z}$ the result is similar to the NLO result, and the scale uncertainties remain largely unchanged. In other words, since $Z+2j$ topologies do not dominate the high- $p_{t,Z}$ distribution, adding NLO corrections to them (i.e. \bar{n} NLO $Z+j$) makes no difference either to the result or to the uncertainties. We have also shown the dependence on the choice of R in the LoopSim procedure. It is smaller

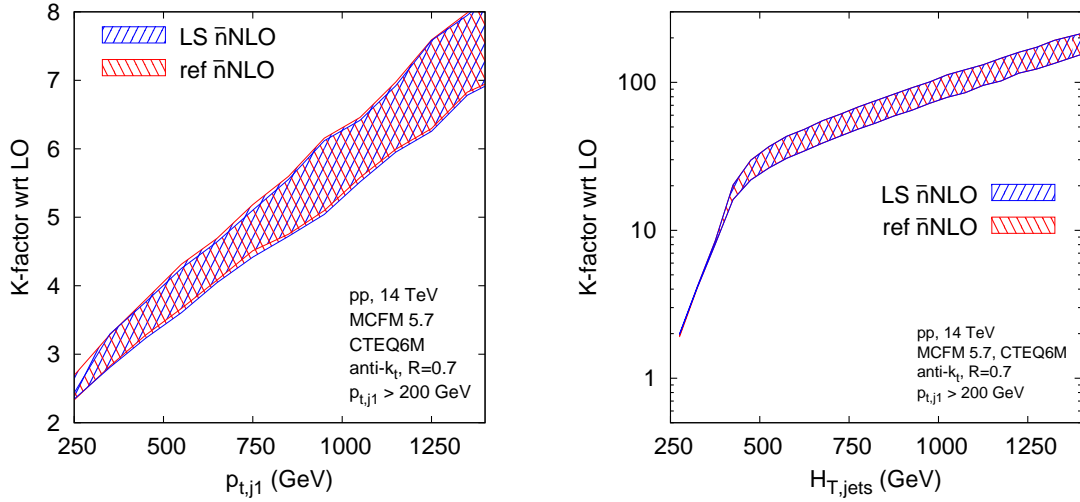


Figure 9: Comparison between the approximate NNLO/LO K -factor calculated using respectively the LoopSim and the “reference-observable” method for $p_{t,j1}$ and $H_{T,jets}$. As a reference observable we have used the differential cross section for $p_{t,Z}$.

than the scale dependence.

The $p_{t,j1}$ distribution gets a correction that is just within the NLO uncertainty band, with \bar{n} NLO uncertainties that are about half the size of the NLO band. Adding in the \bar{n} NLO term has made a real difference. This is precisely what we expect: the observable is dominated by Z+2-parton configurations, and these were only present at tree-level in the NLO Z+j calculation. Our use of \bar{n} NLO provides the additional 1-loop Z+2-parton and tree-level Z+3-parton configurations that come with NLO Z+2j accuracy.

Given the improvement in scale uncertainty, we need to ask whether the uncertainty due to R_{LS} variation might somehow eliminate part of this benefit. It is, however, small. The reasons are similar to those given around eq. (28).

The $H_{T,jets}$ and $H_{T,tot}$ distributions get significant \bar{n} NLO corrections, with \bar{n} NLO/NLO K -factors of about $1.7 - 2$. Absolute scale uncertainties increase slightly compared to NLO, but because of the large K -factor, relative scale uncertainties diminish. At first sight, it is somewhat disturbing that the \bar{n} NLO and NLO uncertainty bands don’t overlap. Given the novelty of the LoopSim method, one should therefore ask whether this is reasonable and whether there is any way of cross-checking the result.

A first observation is that since \bar{n} NLO Z+j is really NLO of the dominant Z+2j component, the large \bar{n} NLO corrections that we see are comparable to an $\mathcal{O}(2)$ K -factor for going from LO to NLO in the Z+2j prediction. There are many contexts where NLO and LO results are not compatible within scale uncertainties, and so it is not unreasonable that the same should be seen here.

Still, we would like to have some more quantitative cross checks that our results are sensible. One option is to consider the alternative “reference-observable” method presented in section 3, which only makes use of standard NLO calculations to compute the approximate NNLO corrections. The comparison between the two methods is shown in fig. 9 for $H_{T,jets}$ and $p_{t,j1}$, where we have taken $p_{t,Z}$ as the reference observable. One notices near perfect agreement for $H_{T,jets}$ and very good agreement for $p_{t,j1}$. This gives us

some degree of confidence that the \bar{n} NLO LoopSim results provide an accurate description of the NNLO behaviour for these observables.

A second option for cross-checking the large \bar{n} NLO effects for $H_{T,\text{jets}}$ and $H_{T,\text{tot}}$, is to examine whether H_T type observables might generally be “difficult”. To do so we look at them in the case of QCD jet events.

6 QCD jet events as a testing ground

We have seen that the \bar{n} NLO K -factors for the two effective-mass variables, $H_{T,\text{tot}}$ and $H_{T,\text{jets}}$, in Z+jet(s) events are about a factor of two above the NLO K -factor.

Since NLO is the first order at which we see the dominant “dijet” topology for the H_T variables in Z+jet(s), fig. 2B,C, it might be instructive to establish a correspondence with a simpler process, QCD dijet production. Having a NLO Z+j prediction is analogous to a LO dijet prediction; and the \bar{n} NLO Z+j predictions should be analogous to NLO dijet predictions. NLO cross sections for dijet observables can be calculated exactly and therefore we can check whether NLO K -factors of order 2 appear for effective-mass observables in pure QCD events.

We will consider several effective-mass observables: an $H_{T,n}$ variable, which sums over the n hardest jets above some threshold ($p_{t,\text{min}} = 40 \text{ GeV}$; such a cut is often imposed experimentally¹⁰)

$$H_{T,n} = \sum_{i \in \text{jets with } p_{t,i} > p_{t,\text{min}}}^n p_{t,i}, \quad (29)$$

where $p_{t,i}$ is the transverse momentum of the i^{th} hardest jet. Upper limits on the number of jets included in the effective mass are common in SUSY searches [3, 4]. We also define an effective mass for all jets above the $p_{t,\text{min}}$ threshold,

$$H_T \equiv H_{T,\infty}, \quad (30)$$

which is similar to the $H_{T,\text{jets}}$ and $H_{T,\text{tot}}$ observables of section 5. Finally, for completeness we will consider the distributions of $p_{t,j1}$, $p_{t,j2}$ and the inclusive jet spectrum. All our results in this section will be for a centre-of-mass energy of 7 TeV, to allow comparison to results in the current run of the LHC.

At LO, the distributions of $\frac{1}{2}H_{T,n}$ ($n \geq 2$), $\frac{1}{2}H_T$, $p_{t,j1}$, and $p_{t,j2}$ will all be identical. The inclusive jet spectrum will have a distribution that is twice as large (because each of the two jets contributes). Note that we do not impose any rapidity acceptance limits on the jets: though such a cut would have been trivial to include in the LoopSim procedure, it would have complicated somewhat the reference-observable approach that we will consider at the end of the section. LoopSim results with a rapidity cuts of $|y| < 2$ on the jets are available from the authors on request.

Figure 10(left) shows the distributions for two observables, $\frac{1}{2}H_T$ and $p_{t,2}$ at LO (where they are identical) and at NLO, as determined using NLOJet++ [42, 43] with CTEQ6M PDFs. A first comment is that H_T receives a NLO K -factor of order 2, just like the \bar{n} NLO

¹⁰In section 5 we did not apply this kind of cut on the H_T variables; one purpose in applying it here is to ascertain whether the large higher-order effects persist even with it.

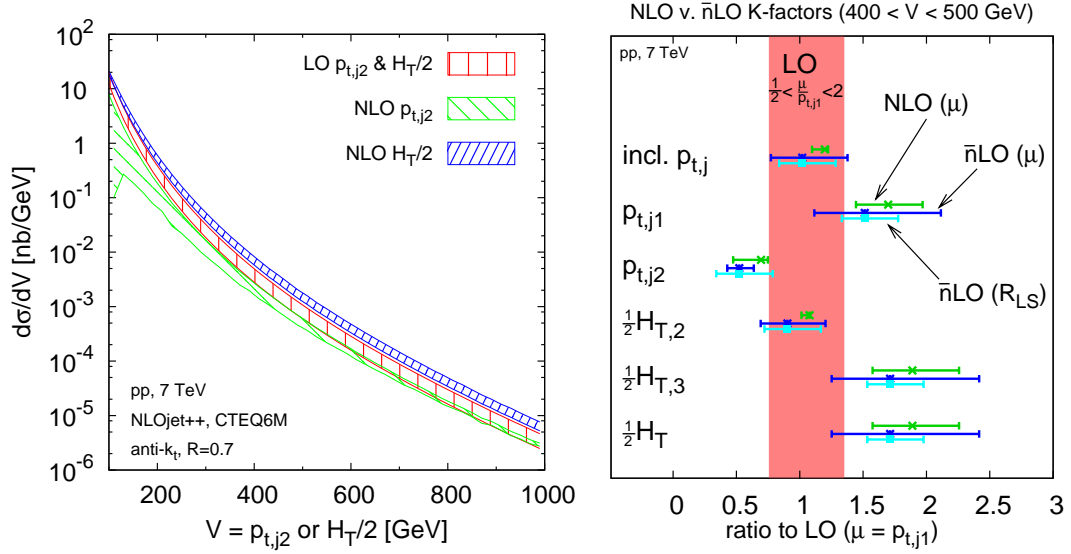


Figure 10: Left: differential cross sections for the $p_{t,j2}$ and $\frac{1}{2}H_T$ observables, at LO, where they are identical, and at NLO where they have substantially different K -factors. Right: the NLO K -factors for the $400 < V/\text{GeV} < 500$ bin for each choice of variable V among the following: the inclusive jet spectrum, the p_t distribution of the hardest ($p_{t,j1}$) and second hardest ($p_{t,j2}$) jets, (half) the effective mass of the two hardest jets ($H_{T,2}$), three hardest jets ($H_{T,3}$) and of all jets above 40 GeV (H_T). Also shown on the right are the \bar{n} LO results for the K -factors. The NLO and \bar{n} LO (μ) widths correspond to the uncertainty due to simultaneous renormalisation and factorisation scale variation by a factor of two around a central value $\mu = p_{t,j1}$. The \bar{n} LO(R_{LS}) width shows the uncertainty from a variation of R_{LS} in the range $0.5 < R_{LS} < 1.5$.

enhancements in the $Z+j$ case. This provides supporting evidence as to their legitimacy. A second comment is that the cross sections are large: these observables will be easily accessible with a few pb^{-1} of integrated luminosity at a 7 TeV LHC, allowing for an early experimental verification of the large K -factor for H_T .

The other observable in the left-hand plot of fig. 10, $p_{t,j2}$, has a very different K -factor, somewhat below 1. The right-hand plot shows the NLO K -factors for our full range of observables, focusing on a single bin of the left-hand one, from 400 – 500 GeV. The pattern that we see here allows us to make some deductions. Firstly, the $H_{T,2}$ variable, which sums the p_t 's of the two leading jets, is free of large NLO enhancements. It is the addition of the third jet in $H_{T,3}$ and H_T that brings about the enhancement. A natural interpretation is the following: it is common for a third, soft jet to be present due to initial state radiation. This third jet shifts the H_T distribution to slightly larger values, and because the distribution falls very steeply, that leads to a non-negligible enhancement. This suggests that if, in section 5, we had used effective mass observables with at most two objects in the sum, then the \bar{n} NLO/NLO ratios would have been close to 1. We have verified that this is indeed the case.

The pattern for $p_{t,1}$ and $p_{t,2}$ in fig. 10 can also be explained in similar terms: a soft ISR emission boosts the hard dijet system, breaking the degeneracy between the p_t 's of the two hardest jets. It is jet 1 that shifts to larger p_t (giving a K -factor > 1), while jet

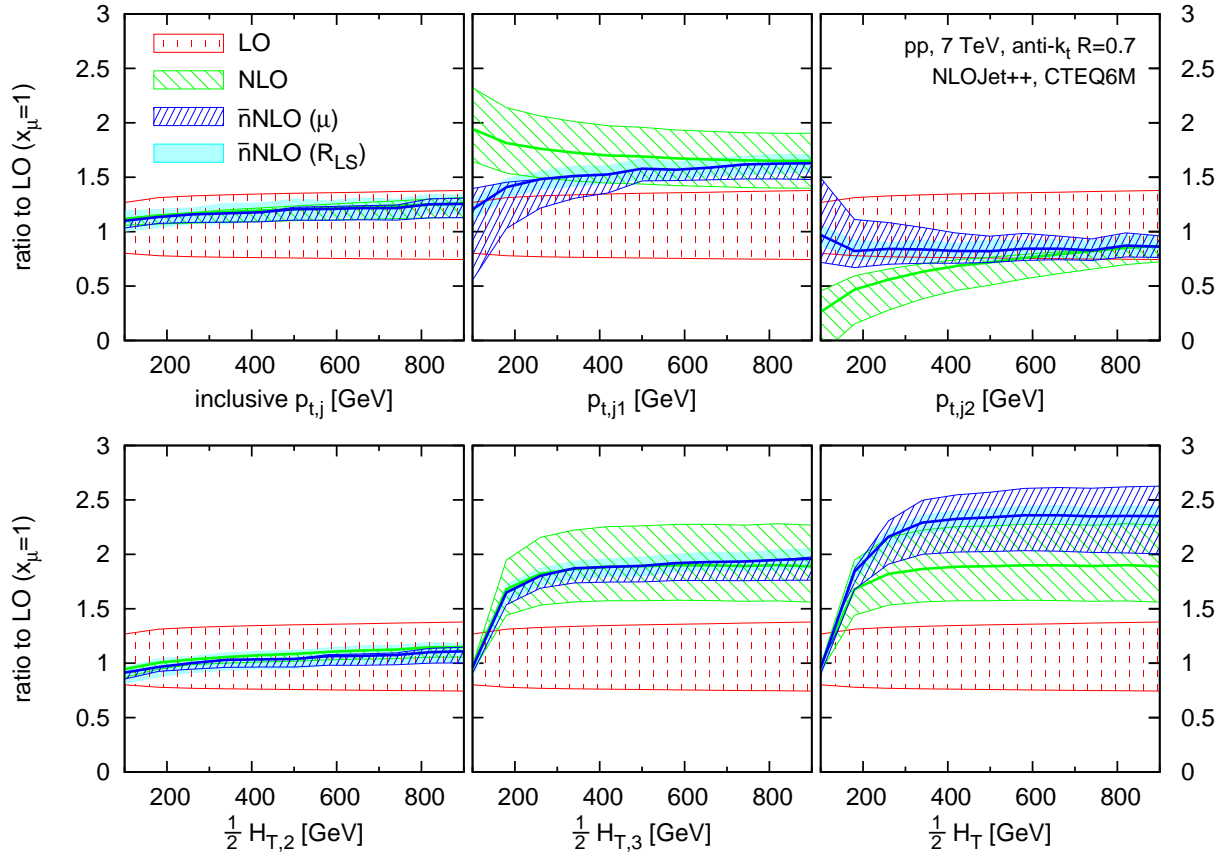


Figure 11: The \bar{n} NLO and NLO K -factors relative to the LO predictions, as a function of p_t (or $\frac{1}{2}H_T$, etc.), for the collection of jet observables described in the text.

2 shifts to lower p_t and so it gets a K factor below 1. For the inclusive jet spectrum, and for $H_{T,2}$, this effect balances out. In addition, final-state radiation from one of the jets can cause it to shift to lower p_t (becoming the 2nd jet), further reducing the K -factor for the distribution of $p_{t,j2}$.

Of the different variables, it is only the inclusive jet p_t and $H_{T,2}$ for which there is a clear reduction in scale uncertainty in going from LO to NLO.

Figure 10(right) also shows the \bar{n} LO results (including uncertainties both from scale variation and from the LoopSim parameter R_{LS}). Despite the fact that none of the K -factors is parametrically large (except arguably for $H_{T,3}$ and H_T), the \bar{n} LO results are remarkably effective at reproducing the pattern of NLO K -factors, albeit with a small systematic shift and generally larger scale uncertainties. One can also verify that, to within 10 – 20%, the p_t dependence of the NLO K -factors is reproduced at \bar{n} LO.

Given this success of \bar{n} LO, and the observed limited convergence of some of the observables at NLO, it is interesting to examine what happens at \bar{n} NLO, where the additional 3j@NLO contribution that we require is again obtained using NLOJet++. Results are shown in fig. 11.

For the inclusive jet spectrum and $H_{T,2}$, which already saw large reductions in scale-dependence at NLO, the \bar{n} NLO corrections have essentially no meaningful effect: they neither significantly affect the central values, nor reduce the scale uncertainties. For these

observables, NLO already converged well, and adding a subset of the NNLO corrections without the 2-loop part cannot improve the result.

For the other effective mass observables, the situation is quite different. With $H_{T,3}$, the \bar{n} NLO result is close to the NLO result and the scale uncertainty is much reduced, i.e. this observable seems to come under control at \bar{n} NLO. In contrast, H_T is subject to quite a large further correction, with the central value at \bar{n} NLO lying outside the NLO uncertainty band, and the \bar{n} NLO uncertainty band (dominated by scale variation) only marginally smaller than at NLO. Why is this? Perhaps we are seeing the effect of a second ISR emission, which shifts the H_T distribution to even higher values? Given that $H_{T,3}$ converges and H_T does not, such an explanation is not unattractive. It is also consistent with the decrease in K -factor at low H_T , where the 40 GeV p_t cutoff on the jets contributing to the H_T sum will eliminate the ISR enhancement. A definitive conclusion would however probably require further study.

For the remaining two observables, $p_{t,1}$ and $p_{t,2}$, the \bar{n} NLO contribution goes in the opposite direction from the NLO correction and at low p_t it seems that the series fails to converge. This is, we believe, closely related to observations of insufficiencies of NLO predictions for dijet cross sections in DIS and photoproduction when identical p_t cuts are imposed on both jets [44, 45, 46, 47, 48] (equivalent to integrating the p_{t2} distribution above that cut). The worse convergence at low p_t is probably due to the larger fraction of subprocesses that involve gluons in the underlying $2 \rightarrow 2$ scattering, so that perturbative corrections tend to go as $(C_A\alpha_s/\pi)^n$ rather than as $(C_F\alpha_s/\pi)^n$ at higher p_t .

Considering that we do not have giant NLO K -factors for the jet processes shown here, one may question the validity of the information obtained from the LoopSim procedure. An important cross check comes from a comparison with the reference-observable technique. Examining fig. 10 (right), one sees two natural reference observables: the inclusive jet spectrum and $H_{T,2}$, both of which show “perturbative” K -factors and small scale dependence at NLO. Here we will use (half) the inclusive jet spectrum as the reference observable (results with $H_{T,2}$ would be almost identical).

Figure 12 provides a comparison of the LoopSim \bar{n} NLO results (showing the envelope of the scale and R_{LS} uncertainties) with the reference-observable \bar{n} NLO results. The comparison is given for all observables except the reference observable itself. The agreement between the two methods is striking, with the reference-observable method giving just a small shift of the K -factors relative to the LoopSim results. The shift is identical for all the observables, as it has to be: it is simply equal to the difference between the NLO and \bar{n} NLO results for the reference observable. Insofar as we believe the scale dependence to be representative of the true NLO uncertainty on the inclusive jet spectrum,¹¹ the results for the other observables should therefore be good approximations to the full NNLO results.

¹¹In light of the fact that the \bar{n} NLO uncertainty for the inclusive jet spectrum is larger than the NLO uncertainty, it may be that our symmetric scale variation is underestimating somewhat the uncertainties present at NLO. To be conservative, it might have been safer to vary the renormalisation and factorisation scales independently.

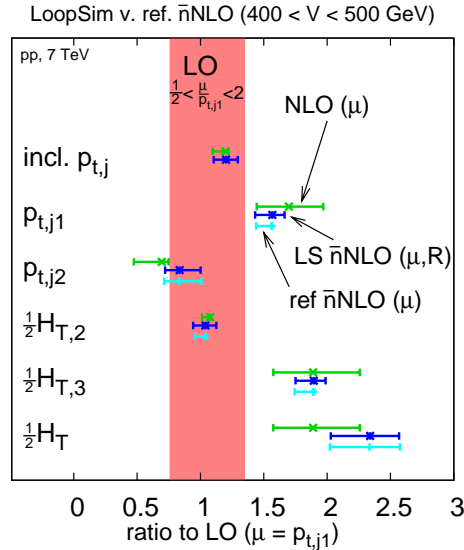


Figure 12: Comparison of LoopSim based \bar{n} NLO results with those from the reference-observable method, here using the inclusive jet p_t spectrum as the reference observable. In the LoopSim results (labelled \bar{n} NLO), the uncertainty bar spans the envelope of the scale and R_{LS} uncertainties. The results are for the $400 < V/\text{GeV} < 500$ bin for each observable V , as in fig. 10.

7 Conclusions

Several cases of LHC observables with giant NLO K -factors have come to light in recent years. They are characterised by the presence at NLO of new partonic scattering topologies that have large enhancements over the LO topologies. In these cases, NLO calculations, while important in highlighting the presence of the large K -factors, cannot on their own provide accurate predictions.

In this article we have examined how to address this problem by combining NLO results for different multiplicities, for example $Z+j@NLO$ with $Z+2j@NLO$. Our main, most flexible method, LoopSim, makes use of unitarity to cancel the infrared and collinear divergences that appear when one tries, say, to apply $Z+2j@NLO$ calculations to observables that are non-zero starting from $Z+1$ -parton. We referred to the result as $Z+j@\bar{n}NLO$, where the “ \bar{n} ” indicates that the highest loop contribution to the NNLO result (the two-loop part) has been estimated with LoopSim.

In introducing a new approximate method for estimating NNLO corrections, significant evidence needs to be provided that the method is meaningful. Firstly, we gave reasons why, in cases with giant K -factors associated with new NLO topologies, we expect $\bar{n}NLO$ results to be a good approximation to NNLO results. As a next step, we carried out studies comparing $Z/\gamma^*@\bar{n}NLO$ (DY) to NNLO predictions for the $pp \rightarrow Z/\gamma^* + X \rightarrow e^+e^- + X$ process. In comparing the DY lepton p_t $\bar{n}NLO$ distributions to NNLO we found near-perfect agreement in a region of giant K -factors, $p_t - \frac{1}{2}m_Z \gtrsim \Gamma_Z$. Interestingly, even in the region where the NLO K -factor was not large, $p_t \lesssim \frac{1}{2}m_Z$, the $\bar{n}NLO$ results provided a significantly better approximation to NNLO than did the plain NLO result. This need not always be the case, but is, we believe, connected to the ob-

servation that our \bar{n} LO results reproduced much of the structure seen at NLO (recall, Z@ \bar{n} LO means combining Z@LO with Z+j@LO).

For Z+j production, the first step of our validation procedure was to compare \bar{n} LO and NLO results. All observables with giant K -factors showed good agreement between the two (one with a moderately large K -factor did not). For those observables, \bar{n} NLO always appeared to provide extra information: either suggesting a convergence of the perturbative series, with reduced scale uncertainties (for $p_{t,j1}$), or an indication of substantial further higher order corrections (for the effective-mass type observables $H_{T,jets}$ and $H_{T,tot}$). Almost identical results were seen with our alternative “reference-observable” estimate of the NNLO contribution.

The large \bar{n} NLO corrections that we saw for effective mass observables led us to examine a range of effective-mass and jet observables in the simpler context of pure jet events (with the expectation that Z+j@NNLO might be similar to 2j@NLO). There we saw a significant NLO K -factor for all effective mass variables except one, $H_{T,2}$, which summed over just the two leading jets. In the Z+j case we had summed over all jets and hence it is not surprising that we should have observed substantial \bar{n} NLO/NLO ratios.

Even though the observables in the pure jets case did not display giant K -factors, the pattern of NLO results was remarkably well reproduced at \bar{n} LO. This encouraged us to study \bar{n} NLO predictions, which provided substantial extra information for several of the observables, with the reference-observable method again giving important cross checks. Since the cross sections for the jet observables are large, these results could easily be tested with early LHC data.

We close this article with a few lines on the relation between LoopSim and other predictive methods. There is a close connection between \bar{n} LO (or $\bar{n}\bar{n}$ LO) and CKKW and MLM [24, 49] matching, since they also both provide ways of combining tree-level results with different multiplicities. Of course CKKW and MLM matching provide an interface with parton showers too, which the LoopSim method does not. On the other hand it is significantly easier to include multiple loop orders into the LoopSim method than it is within matrix-element/parton-showering matching procedures (though work is ongoing in this direction see e.g. [50]).

An interesting cross-check of the LoopSim method will come with the completion of the NNLO calculations for the Z+j and dijet processes. At that point the method could also, for example, be used to merge Z@NNLO with Z+j@NNLO, so as to provide an \bar{n} NNLO prediction for quantities like the Drell-Yan lepton p_t spectrum. The value of the LoopSim method also goes hand-in-hand with progress on 1-loop calculations, especially with the prospect of automated NLO calculations now on the horizon (for example [51, 52, 53]).

Note that the LoopSim code, which will be made public in due course, can currently only deal with hadron-collider processes involving any number of light partons and up to one vector boson. It would benefit from further work to appropriately include heavy quarks and additional bosons.

Acknowledgements

We are grateful to Fabio Maltoni and Simon de Visscher for stimulating discussions and for sharing results with us on the behaviour of MLM type matching for the Z+jet processes, to Giulia Zanderighi for insightful comments and a careful reading of the manuscript, to Massimiliano Grazzini and Giancarlo Ferrera for assistance with DYNNLO, to Matteo Cacciari and Gregory Soyez for FastJet development work that has been useful in implementing the LoopSim procedure and to Philippe Schwemling and Dirk Zerwas for helpful exchanges about H_T definitions. GPS wishes to thank the High Energy Theory group at Rutgers University, the Princeton University Physics Department and the Aspen Center for Physics for hospitality during various stages of this work. This work was supported by the French Agence Nationale de la Recherche, under grant ANR-09-BLAN-0060.

A Recoil procedure

In this appendix, we provide further details on how we perform the recoil of an event when a particle becomes virtual, including the treatment of the decay products of the Z boson. We first examine the simpler case of a particle that makes a loop with the beam, then we show how to deal with a particle that makes a loop with another particle.

A.1 A particle recombines with the beam

Let us assume that particle i_0 makes a loop with the beam. To balance the event moment, we follow the following procedure:

1. For each particle $i \neq i_0$, store its rapidity y_i .
2. Perform a separate longitudinal boost on each particle so as to bring its rapidity to 0 (i.e. get a purely transverse event).
3. Compute

$$E_{tot} = \sum_{i \neq i_0} E_i, \quad (31)$$

where E_i is the energy of particle i in the purely transverse event.

4. Define

$$k = (E = E_{tot}, \vec{p}_t = \vec{p}_{t,i_0}, p_z = 0), \quad (32)$$

and boost all particles into the rest frame of k (so that the total transverse momentum balances).

5. Perform a longitudinal boost on each particle so that it recovers its original rapidity y_i .

For the case where two particles, i_0 and i_1 , are looped with the beam, replace $i \neq i_0$ with $i \neq i_0, i_1$ and in eq. (32) replace \vec{p}_{t,i_0} with $\vec{p}_{t,i_0} + \vec{p}_{t,i_1}$, etc. In the case where the Z decays, for instance into 2 leptons, the procedure is identical except that we apply to the

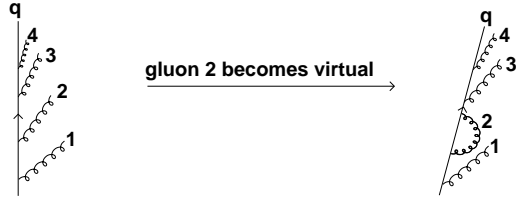


Figure 13: Case where four gluons are emitted from the same quark. Gluon 1 is the last to be clustered with the quark (which roughly corresponds to an early time emission) and gluon 4 is the first to be clustered. In the event where gluon 2 makes a loop over the quark, we spread the gluon 2’s momentum over the quark’s momentum and the momenta of gluons that were emitted *after* it, i.e. gluons 3 and 4 (an earlier time emission like gluon 1 cannot be affected).

leptons the same longitudinal boosts as for the Z (the rapidity of the leptons is thus not necessarily 0 when we apply the transverse boost). This conserves the property that the sum of the leptons’ momenta is still the Z momentum in the “looped” event.

The logic of the above procedure is that if we had attempted to apply a transverse boost without stages 2 and 5, we would have found that our choice of transverse boost, and the corresponding mapping of high- p_t particles’ momenta, would be affected by the presence of energetic particles collinear to the beam. This would have made the procedure collinear unsafe.

A.2 A particle recombines with another particle

Let us consider the situation depicted in fig. 13: four gluons are emitted from the same quark, but at different angles:

$$\theta_{1q} \gg \theta_{2q} \gg \theta_{3q} \gg \theta_{4q}, \quad (33)$$

and gluon 2 becomes virtual. The virtualisation of gluon 2 over the quark cannot have an impact on gluon 1, which was emitted earlier in an angular-ordered picture. But it has an impact on gluons 3 and 4. More precisely, let the p_i be the momenta in the original event and p'_i the momenta in the event where gluon 2 is virtual. We define

$$p_{t,\text{tot}} = p_{t,q} + p_{t,3} + p_{t,4}. \quad (34)$$

and then set the p'_i as follows:

$$p'_i = p_i + \frac{p_{t,i}}{p_{t,\text{tot}}} p_2 \quad \text{for } i = q, 3, 4, \quad (35a)$$

$$p'_1 = p_1. \quad (35b)$$

Subsequently each particle’s p'_i momentum is adjusted such that its mass is 0 (or m_Z if gluon 3 is a Z boson rather than a gluon), keeping its transverse components p_x , p_y and its rapidity unchanged. This can be easily generalised to any number of particles recombining with the same hard one: for each recombined particle i , we spread the looped

particle over the hard particle h and over any non-looped emissions from h that are at smaller angle (i.e. earlier in the C/A clustering sequence) than i . In eqs. (34,35a) it is always the original particle momenta that are used to determine the $p_{t,i}/p_{t,\text{tot}}$ ratio, so that the result is independent of the order in which we perform the recombinations.

This procedure is designed to ensure collinear safety: if, for instance, gluon 4 is collinear to the quark in the original event, then it remains collinear in the looped event. And if it is gluon 4 (emitted after all the others) that is looped, only the quark momentum is rescaled and its direction barely changes, so that angles between the quark and the other gluons stays the same.

In the case where the Z decays into 2 leptons, one applies the following procedure to each of the leptons:

1. Perform a longitudinal boost of the Z boson respectively in the original event and the looped event such that it has 0 rapidity in each case. Call the momenta obtained $p_{Z,0} = (E_0, \vec{p}_{t,0}, 0)$ and $p_{Z,1} = (E_1, \vec{p}_{t,1}, 0)$ respectively.
2. Perform a longitudinal boost of the lepton from the original event into the frame where the initial Z has 0 rapidity.
3. Define a purely transverse vector k such that $p_{Z,0}$ is transformed to $p_{Z,1}$ if it is boosted into k 's rest frame:

$$k = \left(E_1 + E_0, \frac{2}{1+C}(\vec{p}_{t,1} - \vec{p}_{t,0}), 0 \right), \quad (36)$$

with

$$C = \frac{(\vec{p}_{t,1} - \vec{p}_{t,0})^2}{(E_1 + E_0)^2}. \quad (37)$$

4. Boost the lepton's momentum into k 's rest frame.
5. Apply to the lepton the longitudinal boost that brings $p_{Z,1}$ to its true rapidity in the looped event.

We are aware of the cumbersome nature of these procedures. A simplification of them that retained the relevant collinear-safety properties would certainly be of interest.

B The LoopSim method and incoming partons

Without going into a full proof, we shall here illustrate why the LoopSim method is sensible even in the presence of incoming hadrons, by considering what happens at \bar{n} LO. We start with a LO cross section for a process producing n hard objects

$$\sigma_n^{\text{LO}} = \int dx_a dx_b d\Phi_n \frac{d\hat{\sigma}_{ij \rightarrow n}(x_a p_a, x_b p_b)}{d\Phi_n} f_{i/a}(x_a, \mu_F^2) f_{j/b}(x_b, \mu_F^2) C(p_1, \dots, p_n). \quad (38)$$

For compactness of notation, we have dropped the μ_R dependence in the differential tree-level partonic cross section $d\hat{\sigma}_{ij \rightarrow n}/d\Phi_n$. We have also not yet specified our choice for the

factorisation scale μ_F . We assume that $d\hat{\sigma}_{ij \rightarrow n}/d\Phi_n$ contains the necessary constraints to relate the incoming partonic momenta to the outgoing momenta. We further integrate over the phase-space $dx_a dx_b d\Phi_n$, and include a function $C(p_1, \dots, p_n)$, which is 1 if the momenta pass our cuts and 0 otherwise.

We now imagine that there is some transverse-momentum scale Q_0 below which no radiation is emitted. To $\mathcal{O}(\alpha_s)$, the PDFs $f_{i/a}(x_a, \mu_F^2)$ can be written in terms of PDFs at scale Q_0 :

$$f_{i/a}(x_a, \mu_F^2) = f_{i/a}(x_a, Q_0^2) + \frac{\alpha_s}{2\pi} \int_{Q_0^2}^{\mu_F^2} \frac{dk_t^2}{k_t^2} \int \frac{dz}{z} P_{ik}(z) f_{k/a}(x/z, Q_0^2), \quad (39)$$

where we sum implicitly over repeated indices. We also define an unregularised splitting function $p_{ik}(z)$ such that $P_{ik}(z) = p_{ik}(z) - \delta(1-z) \int dz' \bar{p}_{ik}(z')$, with $\bar{p}_{ik}(z')$ embodying the virtual parts of the splitting function (it is zero for $i \neq k$).

Next, we write the LO cross section in terms of a PDF for proton a that has been evaluated at scale Q_0 :

$$\begin{aligned} \sigma_n^{\text{LO}} &= \int dx_a dx_b d\Phi_n \frac{d\hat{\sigma}_{ij \rightarrow n}(x_a p_a, x_b p_b)}{d\Phi_n} f_{j/b}(x_b, \mu_F^2) C(p_1, \dots, p_n) \\ &\times \left[f_{i/a}(x_a, Q_0^2) + \frac{\alpha_s}{2\pi} \int_{Q_0^2}^{\mu_F^2} \frac{dk_t^2}{k_t^2} dz \left(\frac{p_{ik}(z)}{z} f_{k/a}(x_a/z, Q_0^2) - \bar{p}_{ik}(z) f_{i/a}(x_a, Q_0^2) \right) \right]. \quad (40) \end{aligned}$$

Note that the first term in round brackets on the second line corresponds to real emission of a parton. However that parton is not taken into account in the $C(p_1, \dots, p_n)$ factor.

Next we examine the structure of the \bar{n} LO contribution,

$$\begin{aligned} \sigma_n^{\bar{n}\text{LO}} &= \sigma_n^{\text{LO}} + \int dx_a dx_b d\Phi_{n+1} \frac{d\hat{\sigma}_{ij \rightarrow n+1}(x_a p_a, x_b p_b)}{d\Phi_{n+1}} f_{i/a}(x_a, \mu_F^2) f_{j/b}(x_b, \mu_F^2) \\ &\times [C(p_1, \dots, p_{n+1}) - C(p_1^{\text{LS}}, \dots, p_n^{\text{LS}})], \quad (41) \end{aligned}$$

where the $p_1^{\text{LS}} \dots p_n^{\text{LS}}$ represent the momenta when the LoopSim procedure has looped p_{n+1} . In the limit in which p_{n+1} is collinear to incoming parton i , with momentum $p_{n+1} \simeq (1-z)x_a p_a$, the $n+1$ -parton differential cross section and phase-space simplify

$$dx_a d\Phi_{n+1} \frac{d\hat{\sigma}_{ij \rightarrow n+1}(x_a p_a, x_b p_b)}{d\Phi_{n+1}} = dx'_a d\Phi_n \frac{d\hat{\sigma}_{kj \rightarrow n}(x'_a p_a, x_b p_b)}{d\Phi_n} \cdot \frac{\alpha_s}{2\pi} \frac{dz}{z} \frac{dk_{t,n+1}^2}{k_{t,n+1}^2} p_{ki}(z), \quad (42)$$

where $x'_a = zx_a$. By “collinear” we will mean $k_{t,n+1} \ll Q$ where Q is the momentum transfer in the hard process. In this limit we also have that $p_l^{\text{LS}} \simeq p_l$ (for $l \leq n$). So, still working within the collinear limit, we can now rewrite eq. (41) as

$$\begin{aligned} \sigma_n^{\bar{n}\text{LO}} &\simeq \sigma_n^{\text{LO}} + \int dx'_a dx_b d\Phi_n \frac{d\hat{\sigma}_{kj \rightarrow n}(x'_a p_a, x_b p_b)}{d\Phi_n} f_{j/b}(x_b, \mu_F^2) \\ &\times \frac{\alpha_s}{2\pi} \int_{Q_0^2}^{Q^2} \frac{dk_{t,n+1}^2}{k_{t,n+1}^2} \frac{dz}{z} p_{ki}(z) [C(p_1, \dots, p_{n+1}) - C(p_1, \dots, p_n)] f_{i/a}(x'_a/z, \mu_F^2). \quad (43) \end{aligned}$$

Next, we exchange $i \leftrightarrow k$, replace $x'_a \rightarrow x_a$ and change the scale μ_F^2 in $f_{i/a}(x'_a/z, \mu_F^2)$ to be Q_0^2 , which is allowed because it corresponds to an $\mathcal{O}(\alpha_s^2)$ change (while here we consider only $\mathcal{O}(\alpha_s)$):

$$\begin{aligned} \sigma_n^{\bar{n}\text{LO}} &\simeq \sigma_n^{\text{LO}} + \int dx_a dx_b d\Phi_n \frac{d\hat{\sigma}_{ij \rightarrow n}(x_a p_a, x_b p_b)}{d\Phi_n} f_{j/b}(x_b, \mu_F^2) \\ &\times \frac{\alpha_s}{2\pi} \int_{Q_0^2}^{Q^2} \frac{dk_{t,n+1}^2}{k_{t,n+1}^2} \frac{dz}{z} p_{ik}(z) [C(p_1, \dots, p_{n+1}) - C(p_1, \dots, p_n)] f_{i/a}(x_a/z, Q_0^2). \end{aligned} \quad (44)$$

Note now that if we take $\mu_F^2 \sim Q^2$ in eq. (40), then the second term in square brackets in eq. (44) cancels the first term in round brackets in the second line of eq. (40). In other words for initial-state radiation, the action of LoopSim is not so much to provide virtual corrections as to cancel the real-emission terms already included implicitly through the PDFs in the leading order cross section. In contrast, the true virtual terms are already included through the PDFs themselves, i.e. through the second term in round brackets in eq. (40).

As an example, consider $pp \rightarrow Z$. At $\bar{n}\text{LO}$ we will have events such as $gq \rightarrow Zq$, where the outgoing quark comes from collinear initial-state splitting $g \rightarrow q\bar{q}$, with an underlying hard subprocess $\bar{q}q \rightarrow Z$. From these events LoopSim will generate a configuration in which the outgoing quark is “looped”. This will come in with a PDF weight that is the product of a gluon distribution and a quark distribution, so it appears that we have a (negative) $gq \rightarrow Z$ contribution, which would be unphysical. However in the LO cross section with a factorisation scale $\mu_F \sim Q$, when we write $\bar{q}q \rightarrow Z$, part of the \bar{q} PDF comes from $g \rightarrow \bar{q}q$ splitting. If we were just to add the real $gq \rightarrow Zq$ diagram to the LO cross section alone, then in the collinear limit we would be double counting the part already included in the PDF. With the negative “ $gq \rightarrow Z$ ” LoopSim contribution, what happens is that we simply remove the \bar{q} PDF component, generated from $g \rightarrow \bar{q}q$ splitting, that was implicitly included at LO with an incorrect final state (i.e. lacking an outgoing quark), since we are now putting it in with the correct final state through the real $gq \rightarrow Zq$ diagram.

Note that we have not yet worked out the full extension of this discussion to higher orders. The details would depend on the precise higher orders that we have in mind, for example $\bar{n}\bar{n}\text{LO}$ versus $\bar{n}\text{NLO}$. However, regardless of these details, the fundamentally unitary nature of the LoopSim procedure is important in ensuring that the simulated “loops” simply bring about an overall consistent set of final states while maintaining the total cross section as calculated with a sensible factorisation scale choice.

C Secondary emitters in LoopSim

In section 2.1.2 we discussed the special treatment needed for “secondary emitters”, i.e. non-Born particles that have emitted something. In our procedure, secondary emitters do not get looped: when particle j makes a loop over i , this is justified by the collinear enhancement of the matrix element due to i and j being close in angle. But the emission of the same j from the configuration where i is virtual does not have such a collinear enhancement, so one must not take it into account. Another way to understand it is to

consider the example of 2-gluon emission from a $q\bar{q}$ dipole. The squared matrix element for the emission of 2 real energy-ordered ($E_1 \gg E_2$) gluons, g_1, g_2 , can be expressed as [54, 55, 56]

$$M(k_1, k_2) = (4\pi\alpha_s)^2 (C_F^2 W_1 + C_F C_A W_2), \quad (45)$$

with

$$W_1 = 4 \frac{(p_q \cdot p_{\bar{q}})}{(p_q \cdot k_1)(k_1 \cdot p_{\bar{q}})} \frac{(p_q \cdot p_{\bar{q}})}{(p_q \cdot k_2)(k_2 \cdot p_{\bar{q}})}, \quad (46a)$$

$$W_2 = 2 \frac{(p_q \cdot p_{\bar{q}})}{(p_q \cdot k_1)(k_1 \cdot p_{\bar{q}})} \left(\frac{(p_q \cdot k_1)}{(p_q \cdot k_2)(k_2 \cdot k_1)} + \frac{(p_{\bar{q}} \cdot k_1)}{(p_{\bar{q}} \cdot k_2)(k_2 \cdot k_1)} - \frac{(p_q \cdot p_{\bar{q}})}{(p_q \cdot k_2)(k_2 \cdot p_{\bar{q}})} \right). \quad (46b)$$

Since the W_1 term diverges when g_2 is collinear to q or \bar{q} (unlike the W_2 term), it becomes relevant when g_2 is considered to have been emitted from q or \bar{q} independently of g_1 . The W_2 term diverges when g_2 is collinear to g_1 (unlike W_1), so it becomes relevant when g_2 is considered to have been emitted from g_1 . This is depicted in fig. 14, which also shows the virtual corrections (cf. [56]). One notices that the W_2 term only appears when g_1 is real. The diagrams where g_1 is virtual are taken into account when g_2 is emitted from q or \bar{q} . Therefore, g_1 cannot become virtual when g_2 makes a loop over it.

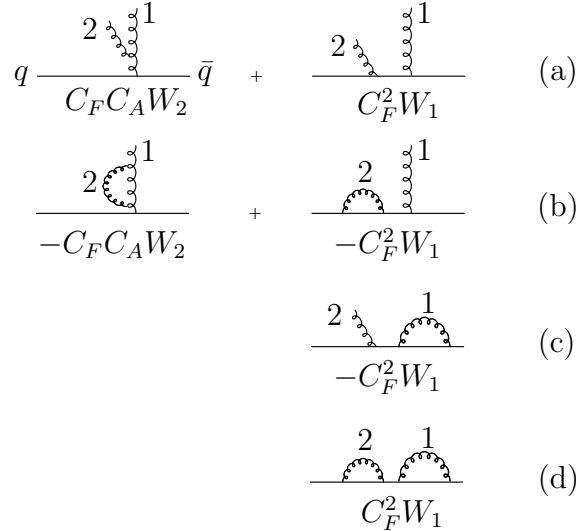


Figure 14: Schematic depiction of the matrix elements for two gluons emitted (or virtual) from a $q\bar{q}$ dipole: (a) gluons 1 and 2 real; (b) gluon 1 real and 2 virtual; (c) gluon 1 virtual and 2 real; (d) gluons 1 and 2 virtual. In each case, when needed, we use the decomposition into W_1 and W_2 pieces to separate what can be seen as the emission of gluon 2 from gluon 1, and what can be seen as the emission of gluon 2 directly from $q\bar{q}$.

References

- [1] I. Hinchliffe, F. E. Paige, M. D. Shapiro, J. Soderqvist and W. Yao, Phys. Rev. D **55**, 5520 (1997) [arXiv:hep-ph/9610544].

- [2] S. Abdullin *et al.* [CMS Collaboration], J. Phys. G **28**, 469 (2002) [arXiv:hep-ph/9806366].
- [3] G. Aad *et al.* [The ATLAS Collaboration], arXiv:0901.0512 [hep-ex].
- [4] G. L. Bayatian *et al.* [CMS Collaboration], J. Phys. G **34**, 995 (2007).
- [5] M. Spiropulu, arXiv:0801.0318 [hep-ex].
- [6] T. Yamazaki [ATLAS Collaboration and CMS Collaboration], arXiv:0805.3883 [hep-ex].
- [7] S. Yamamoto [ATLAS Collaboration], arXiv:0710.3953 [hep-ex].
- [8] M. L. Mangano, Eur. Phys. J. C **59** (2009) 373 [arXiv:0809.1567 [hep-ph]].
- [9] Z. Bern *et al.* [NLO Multileg Working Group], arXiv:0803.0494 [hep-ph].
- [10] J. R. Andersen *et al.* [SM and NLO Multileg Working Group], arXiv:1003.1241 [hep-ph].
- [11] J. M. Campbell, F. Maltoni and F. Tramontano, Phys. Rev. Lett. **98** (2007) 252002 [arXiv:hep-ph/0703113].
- [12] A. Banfi, G. P. Salam and G. Zanderighi, JHEP **0707** (2007) 026 [arXiv:0704.2999 [hep-ph]].
- [13] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100** (2008) 242001 [arXiv:0802.2470 [hep-ph]].
- [14] C. W. Bauer and B. O. Lange, arXiv:0905.4739 [hep-ph].
- [15] A. Denner, S. Dittmaier, T. Kasprzik and A. Muck, arXiv:0906.1656 [hep-ph].
- [16] C. Oleari, presentation at the MC4LHC readiness workshop, CERN, Geneva, March 2010 [<http://indico.cern.ch/getFile.py/access?contribId=46&sessionId=3&resId=0&materialId=slides&confId=74601>].
- [17] F. Febres Cordero, L. Reina and D. Wackerroth, Phys. Rev. D **80** (2009) 034015 [arXiv:0906.1923 [hep-ph]].
- [18] T. Binoth, T. Gleisberg, S. Karg, N. Kauer and G. Sanguinetti, Phys. Lett. B **683** (2010) 154 [arXiv:0911.3181 [Unknown]].
- [19] J. M. Campbell and R. K. Ellis, Phys. Rev. D **60** (1999) 113006 [arXiv:hep-ph/9905386];
J. Campbell, R.K. Ellis, et al. <http://mcfm.fnal.gov>
- [20] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 063 [arXiv:0802.1189 [hep-ph]].
- [21] P. Ciafaloni and D. Comelli, JHEP **0609** (2006) 055 [arXiv:hep-ph/0604070].

- [22] E. W. N. Glover and J. Pires, arXiv:1003.2824 [hep-ph].
- [23] P. Artoisenet, J. M. Campbell, J. P. Lansberg, F. Maltoni and F. Tramontano, Phys. Rev. Lett. **101**, 152001 (2008) [arXiv:0806.3282 [hep-ph]].
- [24] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, JHEP **0111** (2001) 063 [arXiv:hep-ph/0109231].
- [25] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **9708** (1997) 001 [arXiv:hep-ph/9707323].
- [26] M. Wobisch and T. Wengler, arXiv:hep-ph/9907280.
- [27] M. Cacciari and G. P. Salam, Phys. Lett. B **641** (2006) 57 [arXiv:hep-ph/0512210].
- [28] M. Cacciari, G. P. Salam and G. Soyez, <http://fastjet.fr/>.
- [29] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock and B. R. Webber, Phys. Lett. B **269**, 432 (1991);
- [30] S. D. Ellis and D. E. Soper, “Successive Combination Jet Algorithm For Hadron Collisions,” Phys. Rev. D **48**, 3160 (1993) [hep-ph/9305266].
- [31] A. Banfi, G. P. Salam and G. Zanderighi, Eur. Phys. J. C **47**, 113 (2006) [arXiv:hep-ph/0601139].
- [32] S. Hoeche, F. Krauss, S. Schumann and F. Siegert, JHEP **0905** (2009) 053 [arXiv:0903.1219 [hep-ph]].
- [33] S. Frixione, Phys. Lett. B **429** (1998) 369 [arXiv:hep-ph/9801442].
- [34] M. Ciafaloni, P. Ciafaloni and D. Comelli, Phys. Rev. Lett. **84** (2000) 4810 [arXiv:hep-ph/0001142].
- [35] K. Melnikov and F. Petriello, Phys. Rev. D **74** (2006) 114017 [arXiv:hep-ph/0609070].
- [36] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, Phys. Rev. Lett. **103** (2009) 082001 [arXiv:0903.2120 [hep-ph]].
- [37] J. M. Campbell and R. K. Ellis, Phys. Rev. D **62** (2000) 114012 [arXiv:hep-ph/0006304].
- [38] S. Catani and M. Grazzini, Phys. Rev. Lett. **98** (2007) 222002 [arXiv:hep-ph/0703012].
- [39] M. Grazzini, <http://theory.fi.infn.it/grazzini/dy.html>
- [40] J. M. Campbell, R. K. Ellis and D. L. Rainwater, Phys. Rev. D **68** (2003) 094021 [arXiv:hep-ph/0308195].
- [41] C. F. Berger *et al.*, arXiv:1004.1659 [hep-ph].

- [42] Z. Nagy, Phys. Rev. Lett. **88** (2002) 122003. [hep-ph/0110315].
- [43] Z. Nagy, Phys. Rev. **D68** (2003) 094002. [hep-ph/0307268].
- [44] M. Klasen and G. Kramer, Phys. Lett. B **366**, 385 (1996) [hep-ph/9508337].
- [45] S. Frixione and G. Ridolfi, Nucl. Phys. B **507** (1997) 315 [hep-ph/9707345].
- [46] S. Chekanov *et al.* [ZEUS Collaboration], Eur. Phys. J. C **23** (2002) 13 [arXiv:hep-ex/0109029].
- [47] A. Aktas *et al.* [H1 Collaboration], Eur. Phys. J. C **33** (2004) 477 [arXiv:hep-ex/0310019].
- [48] A. Banfi and M. Dasgupta, JHEP **0401** (2004) 027 [arXiv:hep-ph/0312108].
- [49] J. Alwall *et al.*, Eur. Phys. J. C **53** (2008) 473 [arXiv:0706.2569 [hep-ph]].
- [50] N. Lavesson and L. Lonnblad, JHEP **0812** (2008) 070 [arXiv:0811.2912 [hep-ph]].
- [51] R. K. Ellis, K. Melnikov and G. Zanderighi, arXiv:0906.1445 [hep-ph].
- [52] C. F. Berger *et al.*, arXiv:0907.1984 [hep-ph].
- [53] G. Bevilacqua, M. Czakon, C. G. Papadopoulos and M. Worek, arXiv:1002.4009 [hep-ph].
- [54] A. Bassetto, M. Ciafaloni and G. Marchesini, Phys. Rept. **100**, 201 (1983).
- [55] F. Fiorani, G. Marchesini and L. Reina, Nucl. Phys. B **309**, 439 (1988).
- [56] Y. L. Dokshitzer, G. Marchesini and G. Oriani, Nucl. Phys. B **387** (1992) 675.